

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 2, 2004, 21:20:06 ; Search time 1295 Seconds
(without alignments)
6791.946 Million cell updates/sec

Title: US-09-936-506-1
Perfect score: 215
Sequence: 1 ataacggttagctgaggag.....gacagagacgaggaaccgc 215

Scoring table: IDENTITY_NUC
Gapop 10_0, Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues
Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rod.*
- 36: em.htg.mam.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	215	100.0	215	6	AX036038	Sequence
2	215	100.0	215	6	AX036039	Sequence
3	215	100.0	533	6	AR084816	Sequence
4	215	100.0	533	6	I21401	Sequence 2
5	215	100.0	533	6	I95741	Sequence 2
6	215	100.0	2465	6	E12384	Human cDNA
7	215	100.0	2691	6	AR262810	Sequence
8	215	100.0	2691	9	HUMHSP70D	Human heat
9	215	100.0	4360	9	AB018045	Homo sapi
10	214.6	99.8	488	9	S52686	HLA class I
11	214	99.5	59836	9	AL929592	Human DNA
12	214	99.5	125350	2	AC020768	Homo sapi
13	214	99.5	178460	2	AL139040	Homo sapi
14	214	99.5	179894	9	AL662834	Human DNA
15	214	99.5	180283	9	AF134726	Homo sapi
16	213.4	99.3	549	9	HSHSP70A	Human hsp 7
17	212.4	98.8	2700	9	HUMHSP	Human MHC c
18	212.4	98.8	100000	9	AP000503S1	Homo sapi
19	212.4	98.8	113582	9	AL671762	Human DNA
20	210.4	97.9	1903	9	AK096017	Homo sapi
21	203	94.4	214	6	AX036046	Sequence
22	187.4	87.2	420	6	BD108743	EST and e
23	186.2	86.6	511	6	AX194930	Sequence
24	173.4	80.7	2388	9	BC002453	Homo sapi
25	163.6	76.1	2876	9	HUMHSP2	Human MHC c
26	162	75.3	125350	2	AC020768	Homo sapi
27	156.8	72.9	957	9	BC001876	Homo sapi
28	145.2	67.5	9899	6	AX347344	Sequence
29	145.2	67.5	349980	6	AX344568	Sequence
30	134	62.3	9899	6	AX347345	Sequence
31	134	62.3	349980	6	AX344556	Sequence
32	134	62.3	349980	6	AX344557	Sequence
33	109.2	50.8	180	6	AX036044	Sequence
34	109.2	50.8	2465	9	CAHSP70	Sequence
35	74.2	34.5	2445	9	BC018740	Homo sapi
36	56.4	26.2	168	9	HUMHSP70E	Human MHC c
37	36.6	17.0	750	6	AR084815	Sequence
38	36.6	17.0	750	6	I21400	Sequence 1
39	36.6	17.0	750	6	I95740	Sequence 1
40	36.6	17.0	1315	4	BOVHSP70A	Bovine heat
41	36.6	17.0	147172	9	AC022234	Homo sapi
42	36.6	17.0	178622	2	AC026020	Homo sapi
43	36.6	17.0	193333	9	AC090958	Homo sapi
44	36.6	17.0	198264	9	AC018841	Homo sapi
45	35.8	16.7	121394	2	AL137799	Homo sapi

ALIGNMENTS

RESULT 1	AX036038	Sequence 1	215 bp	DNA	linear	PAT 15-NOV-2000
LOCUS	AX036038	Sequence 1	from Patent WO0053785.			
DEFINITION	AX036038					
ACCESSION	AX036038					
VERSION	AX036038.1	GI:11191577				
KEYWORDS						
SOURCE	Homo sapiens (human)					
ORGANISM	Homo sapiens					
REFERENCE	1					
AUTHORS	Coste,H.J. and Ellis,J.H.					
TITLE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
JOURNAL	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
Patent:	WO 0053785-A	1	14-SEP-2000;			

COSTE HERVE JEAN CLEMENT (FR) ; GLAXO GROUP LTD (GB) ; ELLIS
JONATHAN HENRY (GB)
FEATURES
F source
Location/Qualifiers
1..215
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 41 a 71 c 63 g 40 t
ORIGIN

Query Match 100.0%; Score 215; DB 6; Length 215;
Best Local Similarity 100.0%; Pred. No. 9.3e-47;
Matches 215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATACGGGTAGCTGAGAGCTGCTGCGACAGTCCACTACCTTTTTCGAGAGTGAATCCC 60
|||||
Db 1 ATACGGGTAGCTGAGAGCTGCTGCGACAGTCCACTACCTTTTTCGAGAGTGAATCCC 60
|||||

QY 61 GTTGTCCTCAAGCTTCCAGAGCGAACTGTGCGGCTGCGAGCACCGCGGCTCGAGTTT 120
|||||
Db 61 GTTGTCCTCAAGCTTCCAGAGCGAACTGTGCGGCTGCGAGCACCGCGGCTCGAGTTT 120
|||||

QY 121 CCGCGTCCGGAAGACCGAGCTCTTTCGCGGATCCAGTGTTCGGTTTCAGAGCCCA 180
|||||
Db 121 CCGCGTCCGGAAGACCGAGCTCTTTCGCGGATCCAGTGTTCGGTTTCAGAGCCCA 180
|||||

QY 181 TCTCAGAGCCGAGCGACGAGAGCAGGAGGACCGC 215
|||||
Db 181 TCTCAGAGCCGAGCGCGACGAGAGCAGGAGGACCGC 215
|||||

RESULT 2
AX036039
LOCUS AX036039 215 bp mRNA linear PAT 15-NOV-2000
DEFINITION Sequence 2 from Patent WO0053785.
ACCESSION AX036039
VERSION AX036039.1 GI:11191578
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 Coste, H. J. and Ellis, J. H.
AUTHORS
TITLE Expression
JOURNAL Patent: WO 0053785-A 2 14-SEP-2000;
COSTE HERVE JEAN CLEMENT (FR) ; GLAXO GROUP LTD. (GB) ; ELLIS
JONATHAN HENRY (GB)
FEATURES
F source
Location/Qualifiers
1..215
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
BASE COUNT 41 a 71 c 63 g 40 t
ORIGIN

Query Match 100.0%; Score 215; DB 6; Length 215;
Best Local Similarity 100.0%; Pred. No. 9.3e-47;
Matches 215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATACGGGTAGCTGAGAGCTGCTGCGACAGTCCACTACCTTTTTCGAGAGTGAATCCC 60
|||||
Db 1 ATACGGGTAGCTGAGAGCTGCTGCGACAGTCCACTACCTTTTTCGAGAGTGAATCCC 60
|||||

QY 61 GTTGTCCTCAAGCTTCCAGAGCGAACTGTGCGGCTGCGAGCACCGCGGCTCGAGTTT 120
|||||
Db 61 GTTGTCCTCAAGCTTCCAGAGCGAACTGTGCGGCTGCGAGCACCGCGGCTCGAGTTT 120
|||||

QY 121 CCGCGTCCGGAAGACCGAGCTCTTTCGCGGATCCAGTGTTCGGTTTCAGAGCCCA 180
|||||
Db 121 CCGCGTCCGGAAGACCGAGCTCTTTCGCGGATCCAGTGTTCGGTTTCAGAGCCCA 180
|||||

QY 181 TCTCAGAGCCGAGCGCGACGAGAGCAGGAGGACCGC 215
|||||
Db 181 TCTCAGAGCCGAGCGCGACGAGAGCAGGAGGACCGC 215
|||||

COSTE HERVE JEAN CLEMENT (FR) ; GLAXO GROUP LTD (GB) ; ELLIS
JONATHAN HENRY (GB)
FEATURES
F source
Location/Qualifiers
1..533
/organism="unknown"
BASE COUNT 107 a 161 c 170 g 95 t
ORIGIN

Query Match 100.0%; Score 215; DB 6; Length 533;
Best Local Similarity 100.0%; Pred. No. 8.9e-47;
Matches 215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATACGGGTAGCTGAGAGCTGCTGCGACAGTCCACTACCTTTTTCGAGAGTGAATCCC 60
|||||
Db 235 ATACGGGTAGCTGAGAGCTGCTGCGACAGTCCACTACCTTTTTCGAGAGTGAATCCC 294
|||||

QY 61 GTTGTCCTCAAGCTTCCAGAGCGAACTGTGCGGCTGCGAGCACCGCGGCTCGAGTTT 120
|||||
Db 295 GTTGTCCTCAAGCTTCCAGAGCGAACTGTGCGGCTGCGAGCACCGCGGCTCGAGTTT 354
|||||

QY 121 CCGCGTCCGGAAGACCGAGCTCTTTCGCGGATCCAGTGTTCGGTTTCAGAGCCCA 180
|||||
Db 355 CCGCGTCCGGAAGACCGAGCTCTTTCGCGGATCCAGTGTTCGGTTTCAGAGCCCA 414
|||||

QY 181 TCTCAGAGCCGAGCGCGACGAGAGCAGGAGGACCGC 215
|||||
Db 415 TCTCAGAGCCGAGCGCGACGAGAGCAGGAGGACCGC 449
|||||

RESULT 4
I21401
LOCUS I21401 533 bp DNA linear PAT 07-OCT-1996
DEFINITION Sequence 2 from patent US 5521084.
ACCESSION I21401
VERSION I21401.1 GI:1601755
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 533)
AUTHORS Kowalski, J., Gilbert, S. and Zamb, T. J.
TITLE Bovine heat shock promoter and uses thereof
JOURNAL Patent: US 5521084-A 2 28-MAY-1996;
FEATURES
F source
Location/Qualifiers
1..533
/organism="unknown"
BASE COUNT 107 a 161 c 170 g 95 t
ORIGIN

Query Match 100.0%; Score 215; DB 6; Length 533;
Best Local Similarity 100.0%; Pred. No. 8.9e-47;
Matches 215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATACGGGTAGCTGAGAGCTGCTGCGACAGTCCACTACCTTTTTCGAGAGTGAATCCC 60
|||||
Db 235 ATACGGGTAGCTGAGAGCTGCTGCGACAGTCCACTACCTTTTTCGAGAGTGAATCCC 294
|||||

```

QY 61 GTTGTCCCAAGGCTTCCAGAGCGAAGCTGTGCGGCTGCGAGCACCGGCGGTGAGTTT 120
Db 295 GTTGTCCCAAGGCTTCCAGAGCGAAGCTGTGCGGCTGCGAGCACCGGCGGTGAGTTT 354
QY 121 CGGCGTCCGGAAGGACCGAGCTCTTCTCGCGGATCCAGTGTTCGGTTTCCAGCCCCCAA 180
Db 355 CGGCGTCCGGAAGGACCGAGCTCTTCTCGCGGATCCAGTGTTCGGTTTCCAGCCCCCAA 414
QY 181 TCTCAGAGCCGAGCCGACAGAGAGGAGGAAACCGC 215
Db 415 TCTCAGAGCCGAGCCGACAGAGAGGAGGAAACCGC 449

RESULT 5
LOCUS 195741
DEFINITION Sequence 2 from patent US 5733745.
ACCESSION 195741
VERSION 195741.1 GI:3940211
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 533)
AUTHORS Kowalski, J., Gilbert, S. and Zamb, T.J.
TITLE Bovine heat shock promoter and uses thereof
JOURNAL Patent: US 5733745-A 2 31-MAR-1998;
FEATURES
    source
        1..533
        /organism="unknown"
BASE COUNT 107 a 161 c 170 g 95 t
ORIGIN

Query Match 100.0%; Score 215; DB 6; Length 533;
Best Local Similarity 100.0%; Pred. No. 8.9e-47;
Matches 215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATACGGCTAGCTGAGGAGCTGTGCGACAGTCCACTACCTTTTTCGAGAGTGAATCCC 60
Db 235 ATACGGCTAGCTGAGGAGCTGTGCGACAGTCCACTACCTTTTTCGAGAGTGAATCCC 294
QY 61 GTTGTCCCAAGGCTTCCAGAGCGAAGCTGTGCGGCTGCGAGCACCGGCGGTGAGTTT 120
Db 295 GTTGTCCCAAGGCTTCCAGAGCGAAGCTGTGCGGCTGCGAGCACCGGCGGTGAGTTT 354
QY 121 CGGCGTCCGGAAGGACCGAGCTCTTCTCGCGGATCCAGTGTTCGGTTTCCAGCCCCCAA 180
Db 355 CGGCGTCCGGAAGGACCGAGCTCTTCTCGCGGATCCAGTGTTCGGTTTCCAGCCCCCAA 414
QY 181 TCTCAGAGCCGAGCCGACAGAGAGGAGGAAACCGC 215
Db 415 TCTCAGAGCCGAGCCGACAGAGAGGAGGAAACCGC 449

RESULT 6
LOCUS E12384
DEFINITION Human cDNA encoding HSP70.
ACCESSION E12384
VERSION E12384.1 GI:3251217
KEYWORDS JP 1996322577-A/1.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Demura, H., Nomura, K., Shimizu, S., Raari, T.H. and Hisakawa, Y.
TRANSCRIPTIONAL ABNORMALITY OF INTRACELLULAR HSP70MRNA UNDER
SUSTAINED STRESS LOAD CONTAINING HUMAN ACUTENESS AND CHRONICITY AND
ITS APPLICATION (APPLICATION OF BOTH TRANSCRIPTIONAL EXPRESSION OF
NEW HUMAN INTRACELLULAR HSP70MRNA AND TRANSCRIPTIONAL ABNORMALITY
OF HSP70 MRNA)
Patent: JP 1996322577-A 1 10-DEC-1996;

```

```

COMMENT
HOKEN KAGAKU KENKYUSHO:KK
OS Homo sapiens (human)
PN JP 1996322577-A/1
PD 10-DEC-1996
PF 01-JUN-1995 JP 1995158581
PI DEMURA HIROSHI, NOMURA KAORU, SHIMIZU SHOICHI, PI RAARI
TABURIYU HANKINZU,
PI HISAKAWA YOSHIZO
PC C12N15/09, C12P21/02, C12Q1/68;
CC strandedness: Double;
CC topology: Linear;
FH Key
FH source
    1..2465
    /organism="Homo sapiens"
    /tissue type="lymphocyte"
    263..2185
    CDS
    /product="HSP70"
    FT mutation
    join(1..143,306,2465)
    /note="alternative splicing caused by stress".
FEATURES
    source
        1..2465
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
BASE COUNT 553 a 706 c 765 g 441 t
ORIGIN

Query Match 100.0%; Score 215; DB 6; Length 2465;
Best Local Similarity 100.0%; Pred. No. 8.2e-47;
Matches 215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATACGGCTAGCTGAGGAGCTGTGCGACAGTCCACTACCTTTTTCGAGAGTGAATCCC 60
Db 48 ATACGGCTAGCTGAGGAGCTGTGCGACAGTCCACTACCTTTTTCGAGAGTGAATCCC 107
QY 61 GTTGTCCCAAGGCTTCCAGAGCGAAGCTGTGCGGCTGCGAGCACCGGCGGTGAGTTT 120
Db 108 GTTGTCCCAAGGCTTCCAGAGCGAAGCTGTGCGGCTGCGAGCACCGGCGGTGAGTTT 167
QY 121 CGGCGTCCGGAAGGACCGAGCTCTTCTCGCGGATCCAGTGTTCGGTTTCCAGCCCCCAA 180
Db 168 CGGCGTCCGGAAGGACCGAGCTCTTCTCGCGGATCCAGTGTTCGGTTTCCAGCCCCCAA 227
QY 181 TCTCAGAGCCGAGCCGACAGAGAGGAGGAAACCGC 215
Db 228 TCTCAGAGCCGAGCCGACAGAGAGGAGGAAACCGC 262

RESULT 7
LOCUS AR262810
DEFINITION Sequence 1 from patent US 6331388.
ACCESSION AR262810
VERSION AR262810.1 GI:28074512
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2691)
AUTHORS Malkovsky, M. and Wells, A.D.
TITLE Immune response enhancer
JOURNAL Patent: US 6331388-A 1 18-DEC-2001;
FEATURES
    source
        1..2691
        /organism="unknown"
BASE COUNT 600 a 780 c 831 g 480 t
ORIGIN

Query Match 100.0%; Score 215; DB 6; Length 2691;
Best Local Similarity 100.0%; Pred. No. 8.1e-47;
Matches 215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```
Qy 1 ATACCGCTAGCTGAGGAGCTGCTGCAGACAGTCCACTACCTTTTTCGAGAGTGAATCCC 60
    |||
Db 274 ATACCGCTAGCTGAGGAGCTGCTGCAGACAGTCCACTACCTTTTTCGAGAGTGAATCCC 333
    |||
Qy 61 GTTGTCCCAAGCTTCCAGACGCAACTGTGCGGCTGCAGGACCGCGCGTGCAGATT 120
    |||
Db 334 GTTGTCCCAAGCTTCCAGACGCAACTGTGCGGCTGCAGGACCGCGCGTGCAGATT 393
    |||
Qy 121 CCGCGCTCCGAGACGCAACTGTGCGGCTGCAGGACCGCGCGTGCAGATT 180
    |||
Db 394 CCGCGCTCCGAGACGCAACTGTGCGGCTGCAGGACCGCGCGTGCAGATT 453
    |||
Qy 181 TCTCAGACCGCAGCGCAGACAGAGCAGGGAACCGC 215
    |||
Db 454 TCTCAGACCGCAGCGCAGACAGAGCAGGGAACCGC 488
    |||
```

```
RESULT 8
HUMHSP70D          2691 bp  DNA  linear  PRI 08-NOV-1994
LOCUS              Human heat shock protein (hsp 70) gene, complete cds.
DEFINITION
ACCESSION          M11717 M15432
VERSION            M11717.1 GI:184416
KEYWORDS            HSP70 gene; heat shock protein.
SOURCE              Homo sapiens (human)
ORGANISM            Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE           1 (bases 1 to 2691)
AUTHORS             Hunt,C. and Morimoto,R.I.
TITLES              Conserved features of eukaryotic hsp70 genes revealed by comparison
                    with the nucleotide sequence of human hsp70
JOURNAL             Proc. Natl. Acad. Sci. U.S.A. 82 (19), 6455-6459 (1985)
MEDLINE             86016721
PUBMED              3931075
REFERENCE           2 (bases 94 to 293)
AUTHORS             Morgan,W.D., Williams,G.T., Morimoto,R.I., Greene,J., Kingston,R.E.
                    and Tjian,R.
TITLES              Two transcriptional activators, CCAAT-box-binding transcription
                    factor and heat shock transcription factor, interact with a human
                    hsp70 gene promoter
JOURNAL             Mol. Cell. Biol. 7 (3), 1129-1138 (1987)
MEDLINE             87172780
PUBMED              3561411
COMMENT             Original source text: Homo sapiens DNA.
                    (2) revises (1). Sequence revised July 30, 1992.
FEATURES             Location/Qualifiers
                     1..2691
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
                        /map="6p21.3"
                        /product="70 kDa hsp mRNA"
                        /protein_id="AA52697.1"
                        /db_xref="GI:386785"
                     489..2411
                        /gene="HSP70"
                     489..2411
                        /gene="HSP70"
                        /note="70 kDa"
                        /codon_start=1
                        /product="heat shock protein"
                        /protein_id="AA52697.1"
                     489..2411
                        /db_xref="GI:386785"
                     489..2411
                        /db_xref="GDB:G00-120-058"
                     489..2411
                        /translation="MAKAAVGDIGTYSVCVGFQHKVEIIANDQNRTPSYVAF
                        TDERLLGDAKNQVNPQVDFDAKRLIGKFGDPVQSDMKHWPQVINDGDKPK
                        VQVSGKGTAFYEEISSMVLTKMKEIAEYLVPTVAVITVPVFNDSORATKO
                        AGVATGNLVAIINEPTAAIAYGLDRTGKERNVLFDLGGTFDVSILITDGIPE
                        VKATAGTHLGGEDFNLVNFVEEFKHKKDIQNKRAVRLRTACERAKTLLS
                        STQASLEIDFEGIDFTSITRAFELCSDFRSTLEPVEFKALRDKDKAQIHL
                        VLVGGSTRIPKVKLLQDFNFGLNKSINDEAVGYAAVQAAIIMDKSENVDLL
                        LLDVAPLSLGTAGVMTALIKNSTIPTKQITFTTYSNQPGLVIQVYGERAMT
                        KDNLLGRFELSGIPAPGVQIEVTFIDANGILNVTATDKSTGKANKITITNDKGR
```

```
mRNA
gene
CDS
```

```
LSKEEIRMVQAEKYKAEDVORVSAKNALESYAFNMKSAVEDBLKGI SEADK
KKVLDKOEIVISLDANTLAEKDFEHKRKELEQVNCNPIISGLYQAGGPPGFGAQ
GPKGSGSGFTIEVD"
BASE COUNT        600 a 780 c 831 g 480 t
ORIGIN
Query Match       100.0%; Score 215; DB 9; Length 2691;
Best Local Similarity 100.0%; Pred. No. 8.1e-47;
Matches 215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATACCGCTAGCTGAGGAGCTGCTGCAGACAGTCCACTACCTTTTTCGAGAGTGAATCCC 60
    |||
Db 274 ATACCGCTAGCTGAGGAGCTGCTGCAGACAGTCCACTACCTTTTTCGAGAGTGAATCCC 333
    |||
Qy 61 GTTGTCCCAAGCTTCCAGACGCAACTGTGCGGCTGCAGGACCGCGCGTGCAGATT 120
    |||
Db 334 GTTGTCCCAAGCTTCCAGACGCAACTGTGCGGCTGCAGGACCGCGCGTGCAGATT 393
    |||
Qy 121 CCGCGCTCCGAGACGCAACTGTGCGGCTGCAGGACCGCGCGTGCAGATT 180
    |||
Db 394 CCGCGCTCCGAGACGCAACTGTGCGGCTGCAGGACCGCGCGTGCAGATT 453
    |||
Qy 181 TCTCAGACCGCAGCGCAGACAGAGCAGGGAACCGC 215
    |||
Db 454 TCTCAGACCGCAGCGCAGACAGAGCAGGGAACCGC 488
    |||
```

```
RESULT 9
ABO18045          4360 bp  DNA  linear  PRI 14-APR-2000
LOCUS              Homo sapiens HSP70-1 gene for heat shock protein 72, spliced
DEFINITION          variant, partial cds.
ACCESSION           ABO18045
VERSION             ABO18045.1 GI:4691417
KEYWORDS            HSP70-1; heat shock protein 72; HSP70-Hom; alternative splicing.
SOURCE              Homo sapiens (human)
ORGANISM            Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE           1 (sites)
AUTHORS             Shimizu,S., Nomura,K., Ujihara,M. and Demura,H.
TITLES              An additional exon of stress-inducible heat shock protein 70 gene
                    (HSP70-1)
JOURNAL             Biochem. Biophys. Res. Commun. 257 (1), 193-198 (1999)
MEDLINE             98194576
PUBMED              10092532
REFERENCE           2 (bases 1 to 4360)
AUTHORS             Nomura,K. and Shimizu,S.
TITLES              Direct Submission
JOURNAL             Submitted (27-SEP-1998) Kaoru Nomura, Tokyo Women's Medical
                    University, Department of Medicine 2; 8-1 Kawadacho, Shinjyuku-ku,
                    Tokyo 162-8666, Japan (E-mail:nomura7@parkcity.ne.jp,
                    Tel:81-3-3353-8111(ex.39223), Fax:81-3-3357-6475)
COMMENT             Sequence updated (26-Oct-1998).
FEATURES             Location/Qualifiers
                     1..4360
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
                        /chromosome="6"
                        /map="6p21.3"
                        /complement(1..196)
                        /gene="HSP70-Hom"
                        /complement(1..196)
                        /gene="HSP70-Hom"
                        /complement(1..196)
                        /gene="HSP70-1"
                        /complement(1..196)
                        /note="alternative splicing"
                        /number=1
                        /join(2632..2679,3955..4360)
                        /gene="HSP70-1"
                     gene
                     misc_feature
                     gene
                     exon
                     CDS
```

```
/note="spliced variant"
/codon_start=1
/product="heat shock protein 72"
/protein_id="BAA77235.1"
/db_xref="GI:4591418"
/translation="MKHMPFQVINGDKPKVQVSKGETKAPYBEISSMVLTKKEI
AEAYLGPVPTNAVITVPAYFNDQKQATKDAGVLAGLVRLINEPTAAATAYGLDRT
KGERNLIFDILGGGTGFDVSLITIDGIFEVKATAGDTHLGGEDFDNRQ"
3443..3954
/gene="HSP70-1"
/note="alternative splicing
transcription usually starts from exon 2"
/number=2
3955..>4360
/gene="HSP70-1"
/note="alternative splicing"
/number=3
970 a 1242 c 1147 g 1001 t

BASE COUNT
ORIGIN
Query Match 100.0%; Score 215; DB 9; Length 4360;
Best Local Similarity 100.0%; Pred. No. 7,9e-47;
Matches 215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATAACGGCTAGCTGCTGAGGAGCTGCTGCGACAGTCCACTACCTTTTTCGAGAGTGACTCCC 60
Db 3434 ATAACGGCTAGCTGCTGAGGAGCTGCTGCGACAGTCCACTACCTTTTTCGAGAGTGACTCCC 3493
QY 61 GTTGTCCCAAGCTTCCAGAGCGAACCTGTGCGGCTGCGAGCACCGGCGCGTCCAGTTT 120
Db 3494 GTTGTCCCAAGCTTCCAGAGCGAACCTGTGCGGCTGCGAGCACCGGCGCGTCCAGTTT 3553
QY 121 CCGGCGTCCGGAAGGACCGAGCTCTTCTCGCGATCCAGTGTTCGTTTCCAGCCCCCAA 180
Db 3554 CCGGCGTCCGGAAGGACCGAGCTCTTCTCGCGATCCAGTGTTCGTTTCCAGCCCCCAA 3613
QY 181 TCTCAGAGCCGAGCCGACAGAGAGGAGGAAACCGC 215
Db 3614 TCTCAGAGCCGAGCCGACAGAGAGGAGGAAACCGC 3648

RESULT 10
S52686
LOCUS
DEFINITION
HLA class III polymorphic region: HSP70-1=heat shock protein 70 {5'
region} [human, FGF, WT49, WT51, Genomic, 488 nt].
ACCESSION
S52686
VERSION
S52686.1 GI:263578
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1. .488
AUTHORS
Cascino, I., Sorrentino, R. and Tosi, R.
TITLE
Strong genetic association between HLA-DR3 and a polymorphic
variation in the regulatory region of the HSP70-1 gene
JOURNAL
Immunogenetics 37 (3), 177-182 (1993)
MEDLINE
93131289
PUBMED
8420825
REMARK
GenBank staff at the National Library of Medicine created this
entry [NCBI gisbq 122511] from the original journal article.
This sequence comes from Fig. 2.
Region: HLA class III polymorphic region.
COMMENT
Location/Qualifiers
FEATURES
source
1. .488
/organism="Homo sapiens"
/mol_type="Genomic DNA"
/db_xref="taxon:9606"
1. .488
/partial
/gene="HSP70-1"
/note="heat shock protein 70"
```

```
BASE COUNT 98 a 156 c 147 g 85 t 2 others
ORIGIN
Query Match 99.8%; Score 214.6; DB 9; Length 488;
Best Local Similarity 99.5%; Pred. No. 1.1e-46;
Matches 214; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATAACGGCTAGCTGCTGAGGAGCTGCTGCGACAGTCCACTACCTTTTTCGAGAGTGACTCCC 60
Db 274 ATAACGGCTAGCTGCTGAGGAGCTGCTGCGACAGTCCACTACCTTTTTCGAGAGTGACTCCC 333
QY 61 GTTGTCCCAAGCTTCCAGAGCGAACCTGTGCGGCTGCGAGCACCGGCGCGTCCAGTTT 120
Db 334 GTTGTCCCAAGCTTCCAGAGCGAACCTGTGCGGCTGCGAGCACCGGCGCGTCCAGTTT 393
QY 121 CCGGCGTCCGGAAGGACCGAGCTCTTCTCGCGATCCAGTGTTCGTTTCCAGCCCCCAA 180
Db 394 CCGGCGTCCGGAAGGACCGAGCTCTTCTCGCGATCCAGTGTTCGTTTCCAGCCCCCAA 453
QY 181 TCTCAGAGCCGAGCCGACAGAGAGGAGGAAACCGC 215
Db 454 TCTCAGAGCCGAGCCGACAGAGAGGAGGAAACCGC 488

RESULT 11
AL929592
LOCUS
DEFINITION
Human DNA sequence from clone XXbac-147D11 on chromosome 6,
complete sequence.
ACCESSION
AL929592
VERSION
AL929592.8 GI:27848025
KEYWORDS
HTG.
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1. (bases 1 to 59836)
AUTHORS
Tracey, A.
TITLE
Direct Submission
JOURNAL
Submitted (22-JAN-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk
On Jan 22, 2003 this sequence version replaced gi:26985073.
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
-----
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep/
This sequence
was generated from part of bacterial clone contigs constructed by
the MHC Haplotype Consortium and collaborators. Further information
can be found at
http://www.sanger.ac.uk/HGP/Chr6/MHC
XXbac-147D11 is from a DNA-arts QBL human bac library VECTOR:
```

```
FEATURES
  source
    pBelosAC11.
      Location/Qualifiers
        1. 59836
          /organism="Homo sapiens"
          /mol_type="genomic DNA"
          /db_xref="taxon:9606"
          /chromosome="6"
          /clone="XXbac-147D11"
          /clone.lib="DNA-arts-BAC.1-QBL.1"
BASE COUNT 15404 a 14825 c 14944 g 14659 t
ORIGIN
  Query Match 99.5%; Score 214; DB 9; Length 59836;
  Best Local Similarity 100.0%; Pred. No. 1.3e-46;
  Matches 214; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATACGGCTAGCTGAGAGCTGCTGCGACAGTCCACTACCTTTTTCGAGAGTGAATCCC 60
Db 27358 ATACGGCTAGCTGAGAGCTGCTGCGACAGTCCACTACCTTTTTCGAGAGTGAATCCC 27417
QY 61 GTTGTCCCAAGCTTCCAGAGCACTGTGCGGCTGCGACGACCGGCGTGCAGTTT 120
Db 27418 GTTGTCCCAAGCTTCCAGAGCACTGTGCGGCTGCGACGACCGGCGTGCAGTTT 27477
QY 121 CCGGCTCCGGAAGACCGAGCTCTTCGCGGATCCAGTGTTCGTTTCCAGGCCCCCAA 180
Db 27478 CCGGCTCCGGAAGACCGAGCTCTTCGCGGATCCAGTGTTCGTTTCCAGGCCCCCAA 27537
QY 181 TCTCAGCCCGAGCGCGACGAGAGCAGGGAACCG 214
Db 27538 TCTCAGCCCGAGCGCGACGAGAGCAGGGAACCG 27571
RESULT 12
AC020768
LOCUS
DEFINITION Homo sapiens clone RP11-11L21, WORKING DRAFT SEQUENCE, 24 unordered
pieces.
ACCESSION AC020768
VERSION AC020768.3 GI:7341894
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens
ORGANISM Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 125350)
  Birren,B., Linton,L., Nussbaum,C. and Lander,E.
  linear HTG 30-MAR-2000
  Homo sapiens, clone RP11-11L21
  Unpublished
REFERENCE 2 (bases 1 to 125350)
  Birren,B., Linton,L., Nussbaum,C., Lander,E., Abraham,H., Allen,N.,
  Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Bida,F.,
  Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
  Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
  Dearellano,K., Dewar,K., Domino,M., Doyle,M., Feneator,J.,
  Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
  Gardyna,S., Grant,G., Hages,B., Heaford,A., Horton,L.,
  Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
  Landers,T., Lehoczy,J., Levine,R., Liu,C., Liu,G., Locke,K.,
  MacDonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
  McPheeters,R., Meldrim,J., Menues,L., Morrow,J., Naylor,J.,
  Norman,C.H., O'Connor,T., O'Donnell,P., Olivari,T.M., Peterson,K.,
  Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
  Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
  Stojanovic,N., Subramanian,A., Talamas,J., Testfaye,S., Theodore,J.,
  Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
  Zimmer,A. and Zody,M.
  Direct Submission
  Submitted (09-JAN-2000) Whitehead Institute/MIT Center for Genome
  Research, 320 Charles Street, Cambridge, MA 02141, USA
  On Mar 30, 2000 this sequence version replaced gi:6778558.
  All repeats were identified using RepeatMasker:
  Smit, A.P.A. & Green, P. (1996-1997)
  http://ftp.genome.washington.edu/RM/RepeatMasker.html
  ----- Genome Center
  Center: Whitehead Institute/ MIT Center for Genome Research
  Center code: WIBR
  Web site: http://www-seq.wi.mit.edu
  Contact: sequence_submissions@genome.wi.mit.edu
  ----- Project Information
  Center project name: L3221
  Center clone name: 11.L.21
  ----- Summary Statistics
  Sequencing vector: M13; M77815; 100% of reads
  Chemistry: Dye-terminator Big Dye; 100% of reads
  Assembly program: Phrap; version 0.960731
  Consensus quality: 114079 bases at least Q40
  Consensus quality: 119463 bases at least Q30
  Consensus quality: 121683 bases at least Q20
  Insert size: 133000; agarose-fp
  Insert size: 123050; sum-of-contigs
  Quality coverage: 4.3 in Q20 bases; agarose-fp
  Quality coverage: 4.7 in Q20 bases; sum-of-contigs
  -----
  * NOTE: This is a 'working draft' sequence. It currently
  * consists of 24 contigs. The true order of the pieces
  * is not known and their order in this sequence record is
  * arbitrary. Gaps between the contigs are represented as
  * runs of N, but the exact sizes of the gaps are unknown.
  * This record will be updated with the finished sequence
  * as soon as it is available and the accession number will
  * be preserved.
  * 1 1570: contig of 1570 bp in length
  * 1571 1670: gap of 100 bp
  * 1671 3378: contig of 1708 bp in length
  * 3379 3478: gap of 100 bp
  * 3479 4736: contig of 1258 bp in length
  * 4737 4837: gap of 100 bp
  * 4838 6654: contig of 1818 bp in length
  * 6655 6755: gap of 100 bp
  * 6756 8445: contig of 1691 bp in length
  * 8446 8545: gap of 100 bp
  * 8546 10525: contig of 1980 bp in length
  * 10526 10625: gap of 100 bp
  * 10626 12550: contig of 1925 bp in length
  * 12551 12650: gap of 100 bp
  * 12651 16081: contig of 3431 bp in length
  * 16082 16181: gap of 100 bp
  * 16182 19200: contig of 3019 bp in length
  * 19201 19300: gap of 100 bp
  * 19301 21130: contig of 1830 bp in length
  * 21131 21230: gap of 100 bp
  * 21231 23618: contig of 2388 bp in length
  * 23619 23718: gap of 100 bp
  * 23719 26777: contig of 3059 bp in length
  * 26778 26877: gap of 100 bp
  * 26878 29428: contig of 2551 bp in length
  * 29429 29528: gap of 100 bp
  * 29529 32735: contig of 3207 bp in length
  * 32736 32835: gap of 100 bp
  * 32836 37811: contig of 4976 bp in length
  * 37812 37911: gap of 100 bp
  * 37912 42611: contig of 4700 bp in length
  * 42612 42711: gap of 100 bp
  * 42712 49022: contig of 6311 bp in length
  * 49023 49122: gap of 100 bp
  * 49123 56304: contig of 7182 bp in length
  * 56305 56404: gap of 100 bp
  * 56405 63159: contig of 6655 bp in length
  * 63160 63159: gap of 100 bp
  * 63160 70856: contig of 7697 bp in length
  * 70857 70955: gap of 100 bp
  * 70956 79641: contig of 8685 bp in length
  * 79642 79742: gap of 100 bp
  * 79743 90953: contig of 11211 bp in length
  * 90954 91052: gap of 100 bp
  * 91053
```

* 91053 100495: contig of 9443 bp in length
 * 100496 100595: gap of 100 bp
 * 100596 125350: contig of 24755 bp in length.

FEATURES

Location/Qualifiers

1..125350
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clone="RP11-11L21"
 /clone_lib="RPC1-11 Human Male BAC"

1..1570
 /note="assembly_fragment"

misc_feature

1671..3378
 /note="assembly_fragment"

misc_feature

3479..4736
 /note="assembly_fragment"

misc_feature

4837..6654
 /note="assembly_fragment"

misc_feature

6755..8445
 /note="assembly_fragment"

misc_feature

8546..10525
 /note="assembly_fragment"

misc_feature

10626..12550
 /note="assembly_fragment"

misc_feature

12651..16081
 /note="assembly_fragment"

misc_feature

16182..19200
 /note="assembly_fragment"

misc_feature

19301..21130
 /note="assembly_fragment"

misc_feature

clone_end:17
 vector_side:right

misc_feature

21231..23618
 /note="assembly_fragment"

misc_feature

23719..26777
 /note="assembly_fragment"

misc_feature

26878..29428
 /note="assembly_fragment"

misc_feature

29529..32735
 /note="assembly_fragment"

misc_feature

32836..37811
 /note="assembly_fragment"

misc_feature

37912..42611
 /note="assembly_fragment"

misc_feature

42712..49022
 /note="assembly_fragment"

misc_feature

49123..56304
 /note="assembly_fragment"

misc_feature

56405..63059
 /note="assembly_fragment"

misc_feature

63160..70856
 /note="assembly_fragment"

misc_feature

clone_end:SP6
 vector_side:right

misc_feature

70957..79641
 /note="assembly_fragment"

misc_feature

79742..90952
 /note="assembly_fragment"

misc_feature

91053..100495
 /note="assembly_fragment"

misc_feature

100596..125350
 /note="assembly_fragment"

misc_feature

BASE COUNT 29736 a 31948 c 31380 g 29979 t 2307 others

ORIGIN

Query Match 99.5%; Score 214; DB 2; Length 125350;
 Best Local Similarity 100.0%; Pred. No. 1.2e-46;
 Matches 214; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATACGGCTAGCCTGAGGAGCTGCGCAGTCCACTACCTTTTCGAGAGTACTCCC 60

DB 18263 ATACGGCTAGCCTGAGGAGCTGCGCAGTCCACTACCTTTTCGAGAGTACTCCC 18322

QY 61 GTGTCCCAAGGCTTCCAGAGCGAAGCCTGTGCGGTGTCAGGACCGCGCGTTCAGTTT 120
 DB 18323 GTGTCCCAAGGCTTCCAGAGCGAAGCCTGTGCGGTGTCAGGACCGCGCGTTCAGTTT 18382
 QY 121 CCGCGTCCGAAGGACCGAGCTCTTCGCGGATCCAGTGTTCGTTTCCAGCCCCCAA 180
 DB 18383 CCGCGTCCGAAGGACCGAGCTCTTCGCGGATCCAGTGTTCGTTTCCAGCCCCCAA 18442
 QY 181 TCTCAGAGCCGAGCGCAGAGAGCGGGAACCG 214
 DB 18443 TCTCAGAGCCGAGCGCAGAGAGCGGGAACCG 18476

RESULT 13

AL139040

LOCUS

DEFINITION

AL139040 Homo sapiens chromosome 6 clone RP11-400N1, *** SEQUENCING IN

PROGRESS ***, 25 unordered pieces.

ACCESSION

AL139040 GI:9931292

VERSION

HTG; HTGS PHASE1; HTGS_CANCELLED.

KEYWORDS

Homo sapiens (human)

SOURCE

ORGANISM

Homo sapiens

Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 Sims,S.

Direct Submission

Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonerequest@sanger.ac.uk

On Aug 28, 2000 this sequence version replaced gi:9863518.

----- Genome Center

Center: Sanger Centre

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: humquery@sanger.ac.uk

----- Project Information

Center project name: bA400N1

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Sequencing vector: M13; M7815; 0% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 165559 bases at least Q40

Consensus quality: 169956 bases at least Q30

Consensus quality: 172471 bases at least Q20

Insert size: 176060; sum-of-contigs

Quality coverage: 3.31x in Q20 bases; sum-of-contigs Quality

coverage: 3.11x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently

* consists of 25 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1

* 4213: contig of 4213 bp in length

* 4214 4313: gap of 100 bp

* 4314 14555: contig of 10242 bp in length

* 14556 14655: gap of 100 bp

* 14656 17271: contig of 2616 bp in length

* 17272 17371: gap of 100 bp

* 17372 20670: contig of 3299 bp in length

* 20671 20770: gap of 100 bp

* 20771 38215: contig of 17445 bp in length

* 38216 38316: gap of 100 bp

* 38316 42211: contig of 3896 bp in length

* 42212 42311: gap of 100 bp

* 42312 44598: contig of 2287 bp in length


```

SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS     Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE      Tracey, A.
JOURNAL     Direct Submission
COMMENT     Submitted (23-APR-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
            On Feb 14, 2002 this sequence version replaced gi:18491411.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em: EMBL; Sw:
            SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep
            from a CHORI-502 human bac - COX cell line library VECTOR:
            PTARBAC2.1
            This sequence was generated from part of bacterial clone contigs
            constructed by the MHC Haplotype Consortium and collaborators.
            Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6/MHC.

FEATURES             source
     misc_feature     1..179894
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
                        /chromosome="6"
                        /clone="XXBac-40G17"
                        /clone_lib="CHORI-502"
     repeat_region    43951 a 46272 c 46334 g 43337 t
     BASE COUNT      43951 a 46272 c 46334 g 43337 t
     ORIGIN
Query Match      99.5%; Score 214; DB 9; Length 179894;
Best Local Similarity 100.0%; Pred. No. 1.2e-46;
Matches 214; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATAACGGCTAGCCTCAGGAGCTGCTGCAGACAGTCCACTACCTTTTCGAGAGTGACTCCC 60
Db 79426 ATAACGGCTAGCCTCAGGAGCTGCTGCAGACAGTCCACTACCTTTTCGAGAGTGACTCCC 79485

QY 61 GTTGTCCTCAAGCTTCCAGAGCGCACTGTCGGCTCGAGCACCAGCGCGTTCGAGTTT 120
Db 79486 GTTGTCCTCAAGCTTCCAGAGCGCACTGTCGGCTCGAGCACCAGCGCGTTCGAGTTT 79545

QY 121 CCGGCTCGGAAGACCGAGCTCTTCGCGGATCCAGTGTTCGGTTTCAGGCCCCCAA 180
Db 79546 CCGGCTCGGAAGACCGAGCTCTTCGCGGATCCAGTGTTCGGTTTCAGGCCCCCAA 79605

QY 181 TCTCAGAGCCGCGCAGAGAGCAGGACCG 214
Db 79606 TCTCAGAGCCGCGCAGAGAGCAGGACCG 79639

RESULT 15
LOCUS      AF134726/c
DEFINITION Homo sapiens BAC clone 215012 NG35, NG36, G9A, NG22, G9, HSP70-2,
            HSP70-1, HSP70-HOM, snRNP, G7A, NG37, NG23, and MutsH5 genes,
            complete cds.
ACCESSION   AF134726
VERSION     AF134726.1
KEYWORDS    GI:4529886
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS     Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE      Rowen, L., Qin, S., Madan, A., Dickhoff, R., Dors, M., Madan, A.,
            Hicks, P., Loretz, C., Ratcliffe, A., Abbasi, N., Shaffer, T. and
            Hood, L.
JOURNAL     Sequence of the human major histocompatibility complex class III
REFERENCE   1 (bases 1 to 180283)
AUTHORS     Rowen, L., Qin, S., Madan, A., Dickhoff, R., Dors, M., Madan, A.,
            Hicks, P., Loretz, C., Ratcliffe, A., Abbasi, N., Shaffer, T. and
            Hood, L.
TITLE      Sequence of the human major histocompatibility complex class III
JOURNAL     Submitted (15-MAR-1999) Department of Molecular Biotechnology, Box
REFERENCE   2 (bases 1 to 180283)
AUTHORS     Rowen, L.
JOURNAL     Direct Submission
COMMENT     Submitted (15-MAR-1999) Department of Molecular Biotechnology, Box
            357730 University of Washington, Seattle, WA 98195, USA
            Sequencing methodology: high redundancy shotgun in plasmids.
            Interspersed Repeats were identified with RepeatMasker (available
            from http://ftp.genome.washington.edu/RM/RepeatMasker.html). This
            sequence overlaps cosmid S22A (AF019413) by 2363 bases and BAC
            210G24 (AF129756) by 12177 bases.
FEATURES             source
     misc_feature     1..180283
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
                        /chromosome="6"
                        /map="6p21.3"
                        /clone="BAC 215012"
                        /clone_lib="RBC111"
     repeat_region    1..2363
                        /note="Overlap span with cosmid S22A in AF019413."
                        /complement(103..402)
                        /rpt_family="AluSx"
                        /rpt_type=dispersed
     variation        408..409
                        /note="cosmid S22A: ttct; BAC 215012: tt"
                        /replace="ttct"
     repeat_region    complement(420..480)
                        /rpt_family="L1PA16"
                        /rpt_type=dispersed
     repeat_region    complement(458..732)
                        /rpt_family="L1PA13"
                        /rpt_type=dispersed
     repeat_region    complement(834..1134)
                        /rpt_family="AluSx"
                        /rpt_type=dispersed
     variation        1134..1136
                        /note="cosmid S22A: ct; BAC 215012: ctt"
                        /replace="ct"
     repeat_region    complement(1139..1257)
                        /rpt_family="AluSg1"
                        /rpt_type=dispersed
     variation        1445..1446
                        /note="cosmid S22A: ctt; BAC 215012: ct"
                        /replace="ctt"
     repeat_region    complement(1446..1746)
                        /rpt_family="AluSx"
                        /rpt_type=dispersed
     variation        1843
                        /note="cosmid S22A: t; BAC 215012: c"
                        /replace="t"
     repeat_region    2430..2489
                        /rpt_family="MLT1A2"
                        /rpt_type=dispersed
     repeat_region    complement(2831..2884)
                        /rpt_family="LINE2"
                        /rpt_type=dispersed
     repeat_region    complement(3091..3392)

```

```
/rpt_family="AluY"
/rpt_type=dispersed
repeat_region      complement(3416..3712)
/rpt_family="AluSg"
/rpt_type=dispersed
repeat_region      complement(3719..3790)
/rpt_family="MLT2FA"
/rpt_type=dispersed
repeat_region      3813..4084
/rpt_family="AluSg"
/rpt_type=dispersed
repeat_region      4086..4375
/rpt_family="AluSg"
/rpt_type=dispersed
repeat_region      complement(4417..4729)
/rpt_family="AluSg"
/rpt_type=dispersed
repeat_region      4781..4910
/rpt_family="FLAM A"
/rpt_type=dispersed
repeat_region      complement(4939..4968)
/rpt_family="MLT2E"
/rpt_type=dispersed
repeat_region      complement(4980..5112)
/rpt_family="AluJb"
/rpt_type=dispersed
repeat_region      complement(5167..5468)
/rpt_family="AluSx"
/rpt_type=dispersed
repeat_region      complement(5607..5738)
/rpt_family="LINE2"
/rpt_type=dispersed
repeat_region      5786..5873
/rpt_family="L1ME3A"
/rpt_type=dispersed
repeat_region      5885..5938
/rpt_family="MERSA"
/rpt_type=dispersed
repeat_region      5955..6090
/rpt_family="FLAM C"
/rpt_type=dispersed
repeat_region      6099..6389
/rpt_family="AluSx"
/rpt_type=dispersed
repeat_region      6403..6446
/rpt_family="MERSA"
/rpt_type=dispersed
repeat_region      complement(6711..7011)
/rpt_family="AluSx"
/rpt_type=dispersed
repeat_region      complement(7029..7330)
/rpt_family="AluSx"
/rpt_type=dispersed
repeat_region      7486..7688
/rpt_family="AluJb"
/rpt_type=dispersed
repeat_region      7689..7851
/rpt_family="L1MB8"
/rpt_type=dispersed
repeat_region      7852..8150
/rpt_family="AluSg"
/rpt_type=dispersed
repeat_region      8151..8263
/rpt_family="L1MB6"
/rpt_type=dispersed
repeat_region      8290..8425
/rpt_family="FLAM C"
/rpt_type=dispersed
repeat_region      8451..8730
/rpt_family="AluJb"
/rpt_type=dispersed
repeat_region      8737..8839
/rpt_family="L1MB5"
```

```
repeat_region      /rpt_type=dispersed
8816..8972
/rpt_family="L1ME3"
/rpt_type=dispersed
repeat_region      complement(9034..9357)
/rpt_family="TIGER1"
/rpt_type=dispersed
repeat_region      complement(9358..9659)
/rpt_family="AluSx"
/rpt_type=dispersed
repeat_region      complement(9662..9735)
/rpt_family="TIGER1"
/rpt_type=dispersed
repeat_region      9736..9882
/rpt_family="FLAM C"
/rpt_type=dispersed
repeat_region      9884..10185
/rpt_family="AluSg"
/rpt_type=dispersed
repeat_region      10189..10351
/rpt_family="AluSc"
/rpt_type=dispersed
repeat_region      10357..10419
/rpt_family="AluJb"
/rpt_type=dispersed
repeat_region      complement(10427..10711)
/rpt_type=dispersed
repeat_region      complement(10871..11176)
/rpt_family="AluSg"
/rpt_type=dispersed
repeat_region      complement(12410..12541)
/rpt_family="FLAM C"
/rpt_type=dispersed
repeat_region      complement(12577..12873)
/rpt_family="AluJo"
/rpt_type=dispersed
repeat_region      13413..13701
/rpt_family="AluSg"
/rpt_type=dispersed
repeat_region      13702..13744
/rpt_family="AT rich"
/rpt_type=dispersed
repeat_region      complement(13907..13948)
/rpt_family="MIR"
/rpt_type=dispersed
repeat_region      complement(13995..14046)
/rpt_family="MER57B"
/rpt_type=dispersed
repeat_region      complement(14378..14674)
/rpt_family="AluSx"

Query Match      99.5%; Score 214; DB 9; Length 180283;
Best Local Similarity 100.0%; Pred.No.1.2e-46;
Matches 214; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1      ATACGGCTAGCCTGAGGAGCTGCTGCAGAGTCCACTACCTTTTTCGAGAGTGACTCCC 60
Db      102627  ATACGGCTAGCCTGAGGAGCTGCTGCAGAGTCCACTACCTTTTTCGAGAGTGACTCCC 102568

Qy      61      GTTGTCCCAAGGCTTCCAGAGCGAACCTGTGCGGCTGCAGGACCCGCGCGTTCGAGTTT 120
Db      102567  GTTGTCCCAAGGCTTCCAGAGCGAACCTGTGCGGCTGCAGGACCCGCGCGTTCGAGTTT 102508

Qy      121     CCGCGTCCGAGAGCGAGCTTCTTCGCGGATCCAGTGTTCGGTTTCAGGCCCCCAA 180
Db      102507  CCGCGTCCGAGAGCGAGCTTCTTCGCGGATCCAGTGTTCGGTTTCAGGCCCCCAA 102448

Qy      181     TCTCAGAGCGAGCCGACAGAGACGAGCAACCG 214
Db      102447  TCTCAGAGCGAGCCGACAGAGACGAGCAACCG 102414

Search completed: February 3, 2004, 03:00:35
```

Job time : 1301 secs



THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 2, 2004, 21:15:18 ; Search time 176 Seconds
(without alignments)
3297.609 Million cell updates/sec

Title: US-09-936-506-1

Perfect score: 215

Sequence: 1 ataacggttagctgaggag.....Gacagagcaggggaaccgc 215

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq 19Jun03.*

- 1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
- 2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
- 3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
- 4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
- 5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
- 6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
- 7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
- 8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
- 9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
- 10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
- 11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
- 12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
- 13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
- 14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
- 15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
- 16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
- 17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
- 18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
- 19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
- 20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
- 21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
- 22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
- 23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
- 24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
- 25: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	215	100.0	215	21	Human HSP70A 5' un
2	215	100.0	215	21	Human HSP70A 5' un
3	215	100.0	2691	24	Human cDNA differe
4	215	100.0	2691	24	Human hsp72 encodi
5	213.4	99.3	2465	18	Human heat shock p
6	212.4	98.8	2700	21	Human Hsp72 (heat
7	208.6	97.0	2465	21	Human heat shock p
8	186.8	86.9	2302	21	Human heat shock p

C	9	186.2	86.6	511	23	AAH89560	CNS disorder-relat
C	10	185.2	86.1	2302	21	AAA15621	Human heat shock p
C	11	145.2	67.5	595	24	ABQ30838	Oligonucleotide fo
C	12	145.2	67.5	595	24	ABQ30839	Oligonucleotide fo
C	13	145.2	67.5	596	24	ABQ34140	Oligonucleotide fo
C	14	145.2	67.5	596	24	ABQ34141	Oligonucleotide fo
C	15	145.2	67.5	9899	24	ABL34442	Human immune syste
C	16	134	62.3	595	24	ABQ30840	Oligonucleotide fo
C	17	134	62.3	595	24	ABQ30841	Oligonucleotide fo
C	18	134	62.3	596	24	ABQ34138	Oligonucleotide fo
C	19	134	62.3	596	24	ABQ34139	Oligonucleotide fo
C	20	134	62.3	9899	24	ABL34443	Human immune syste
C	21	109.2	50.8	180	21	AAA94745	African green monk
C	22	103.4	48.1	583	24	ABQ30834	Oligonucleotide fo
C	23	103.4	48.1	583	24	ABQ30835	Oligonucleotide fo
C	24	99.6	46.3	583	24	ABQ30836	Oligonucleotide fo
C	25	99.6	46.3	583	24	ABQ30837	Oligonucleotide fo
C	26	89.2	41.5	584	24	ABQ34144	Oligonucleotide fo
C	27	89.2	41.5	584	24	ABQ34145	Oligonucleotide fo
C	28	83.8	39.0	584	24	ABQ34142	Oligonucleotide fo
C	29	83.8	39.0	584	24	ABQ34143	Oligonucleotide fo
C	30	50	23.3	51	22	AAI28268	Human SNP oligonuc
C	31	38.4	17.9	2636	25	ABX62979	Human activated T
C	32	36.6	17.0	750	15	AAQ68581	5' upstream region
C	33	32.4	15.1	205	25	ABX88007	Corn ear-derived p
C	34	32.2	15.0	7710	22	AAF44662	Novel protein kina
C	35	32.2	15.0	7789	24	AAQ30565	Human kinase poly
C	36	32	14.9	3264	24	ABA93874	Human G protein-co
C	37	31.8	14.8	4722	19	AAV32836	Rabbit low density
C	38	31.8	14.8	4722	22	AAH26491	Rabbit low density
C	39	31.6	14.7	391	24	ABN96266	Gene #2764 used to
C	40	31.6	14.7	1196	14	AAQ34594	Clone for TGF-beta
C	41	30.8	14.3	574	21	AACT5371	Human ORFX ORF926
C	42	30.8	14.3	574	24	AACT5371	Human ORFX poly
C	43	30.4	14.1	694	23	AAQ30548	DNA encoding novel
C	44	30.4	14.1	843	24	ABQ90348	M. capsulatus gene
C	45	30.4	14.1	843	24	ABQ90371	M. capsulatus gene

ALIGNMENTS

RESULT 1

AAA94740

ID AAA94740 standard; DNA; 215 BP.

XX

AC AAA94740;

XX

DT 19-JAN-2001 (first entry)

XX

DE Human HSP70A 5' untranslated region DNA sequence.

XX

KW Human; HSP70A; 5' UTR; untranslated region; heat shock protein;

KW translation efficiency; vaccine; bacterial; viral; parasitic infection;

KW immune-related disease; contraceptive; gene therapy; cancer;

KW cardiovascular disorder; cystic fibrosis; ds.

OS Homo sapiens.

XX

PN WO200053785-A2.

XX

PD 14-SEP-2000.

XX

PF 09-MAR-2000; 2000WO-EP02031.

XX

PR 11-MAR-1999; 99GB-0005498.

XX

PA (GLAX) GLAXO GROUP LTD.

XX

PI Coste HJ, Ellis JH;

XX

DR WPI; 2000-594331/56.

XX

PT Human heat shock protein 5' untranslated region (UTR) transcribed to
 PT provide an RNA molecule having UTR that increases translation
 PT efficiency of polypeptides, useful for treating deficiency in
 PT expression of the polypeptide -
 XX
 PS Claim 7; Page 28; 44pp; English.
 XX
 CC The present sequence is the 5' untranslated region (UTR) of human heat
 CC shock protein (Hsp) 70A gene. This sequence has a high potential to form
 CC secondary structures. This sequence can be used to increase the
 CC translation efficiency of a polypeptide. The present sequence may be
 CC useful in therapeutic or prophylactic vaccination for preventing
 CC bacterial, viral and parasitic infections and also for treating
 CC immune-related diseases and for contraceptive purposes. In addition, the
 CC present sequence may be useful in gene therapy of various disorders such
 CC as cancer, cardiovascular disorders and cystic fibrosis.
 XX
 SQ Sequence 215 BP; 41 A; 71 C; 63 G; 40 T; 0 other;
 Query Match 100.0%; Score 215; DB 21; Length 215;
 Best Local Similarity 100.0%; Pred. No. 3.7e-55;
 Matches 215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 ATACGGCTAGCTGAGGAGCTGCTGCGACAGTCCACTACCTTTTCGAGAGTGACTCCC 60
 Db 1 ATACGGCTAGCTGAGGAGCTGCTGCGACAGTCCACTACCTTTTCGAGAGTGACTCCC 60
 Qy 61 GTTGTCCCAAGGCTTCCAGAGCGAACCTGTGCGGCTCGAGCACCGCGCGTTCGAGTTT 120
 Db 61 GTTGTCCCAAGGCTTCCAGAGCGAACCTGTGCGGCTCGAGCACCGCGCGTTCGAGTTT 120
 Qy 121 CCGGCGTCCGGAAGACCGAGCTCTTCTCGGGATCCAGTGTTCGGTTTCCAGCCCCCAA 180
 Db 121 CCGGCGTCCGGAAGACCGAGCTCTTCTCGGGATCCAGTGTTCGGTTTCCAGCCCCCAA 180
 Qy 181 TCTCAGAGCCGAGCCGACAGAGCAGGGAACCGC 215
 Db 181 TCTCAGAGCCGAGCCGACAGAGCAGGGAACCGC 215
 RESULT 2
 AAA94756
 ID AAA94756 standard; mRNA; 215 BP.
 XX
 AC AAA94756;
 XX
 DT 19-JAN-2001 (first entry)
 XX
 DE Human HSP70A 5' untranslated region mRNA sequence.
 XX
 KW Human; HSP70A; 5' UTR; untranslated region; heat shock protein;
 KW translation efficiency; vaccine; bacterial; viral; parasitic infection;
 KW immune-related disease; contraceptive; gene therapy; cancer;
 KW cardiovascular disorder; cystic fibrosis; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200053785-A2.
 XX
 PD 14-SEP-2000.
 XX
 PF 09-MAR-2000; 2000WO-EP02031.
 XX
 PR 11-MAR-1999; 99GB-0005498.
 XX
 PA (GLAXO) GLAXO GROUP LTD.
 XX
 PI Coste HJ, Ellis JH;
 XX
 DR WPI; 2000-594331/56.
 XX
 PT Human heat shock protein 5' untranslated region (UTR) transcribed to
 PT provide an RNA molecule having UTR that increases translation

PT efficiency of polypeptides, useful for treating deficiency in
 PT expression of the polypeptide -
 XX
 PS Disclosure; Page 3; 44pp; English.
 XX
 CC The present sequence is the 5' untranslated region (UTR) mRNA of human
 CC heat shock protein (Hsp) 70A gene. This sequence has a high potential to
 CC form secondary structures. This sequence can be used to increase the
 CC translation efficiency of a polypeptide. The present sequence may be
 CC useful in therapeutic or prophylactic vaccination for preventing
 CC bacterial, viral and parasitic infections and also for treating
 CC immune-related diseases and for contraceptive purposes. In addition, the
 CC present sequence may be useful in gene therapy of various disorders such
 CC as cancer, cardiovascular disorders and cystic fibrosis.
 XX
 SQ Sequence 215 BP; 41 A; 71 C; 63 G; 40 U; 0 other;
 Query Match 100.0%; Score 215; DB 21; Length 215;
 Best Local Similarity 81.4%; Pred. No. 3.7e-55;
 Matches 175; Conservative 40; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 ATACGGCTAGCTGAGGAGCTGCTGCGACAGTCCACTACCTTTTCGAGAGTGACTCCC 60
 Db 1 AUAACGGCUAGCCUGAGGAGCUGCUGCGACAGUCACUCCUUUUUCGAGAGUCACCCC 60
 Qy 61 GTTGTCCCAAGGCTTCCAGAGCGAACCTGTGCGGCTCGAGCACCGCGCGTTCGAGTTT 120
 Db 61 GUUUGUCCNAGGCUUCCNAGAGCGAACCUUGUGCGGTCGAGCACCGCGCGUGAGUUU 120
 Qy 121 CCGGCGTCCGGAAGACCGAGCTCTTCTCGCGATCCAGTGTTCGGTTTCCAGCCCCCAA 180
 Db 121 CCGGCGUCCGGAAGACCGAGCUCUCUCGCGGAUCCAGUGUCCUUUCCAGCCCCCAA 180
 Qy 181 TCTCAGAGCCGAGCCGACAGAGCAGGGAACCGC 215
 Db 181 UCUCAGAGCCGAGCCGACAGAGCAGGGAACCGC 215
 RESULT 3
 ABK83917
 ID ABK83917 standard; cDNA; 2691 BP.
 XX
 AC ABK83917;
 XX
 DT 14-AUG-2002 (first entry)
 XX
 DE Human cDNA differentially expressed in granulocytic cells #488.
 XX
 KW Human; ss; granulocytic cell; DNA chip; bacterial infection;
 KW viral infection; parasitic infection; protozoal infection;
 KW fungal infection; sterile inflammatory disease; psoriasis;
 KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
 KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
 KW adult respiratory distress syndrome; inflammatory bowel disease;
 KW Crohn's disease; ulcerative colitis; periodontal disease;
 KW granulocyte activation; chronic inflammation; allergy.
 XX
 OS Homo sapiens.
 XX
 PN WO200228999-A2.
 XX
 PD 11-APR-2002.
 XX
 PF 03-OCT-2001; 2001WO-US30821.
 XX
 PR 03-OCT-2000; 2000US-237189P.
 XX
 PA (GENE-) GENE LOGIC INC.
 XX
 PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
 XX
 DR WPI; 2002-435328/46.
 XX

PT Detecting granulocyte activation by detecting differential expression
 PT of genes associated with granulocyte activation, which serves as
 PT diagnostic markers that is useful for monitoring disease states and
 PT drug toxicity

XX Claim 1; SEQ ID No 488; 114pp; English.

CC The invention relates to detecting (M1) granulocyte (GC) activation
 CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
 CC DNA chip analysis as given in the specification, and comparing
 CC the expression level to an expression level in an unactivated
 CC GC, where differential expression of Gs is indicative of GCA.
 CC Also included are modulating (M2) Gs by contacting GC with an agent
 CC that alters the expression of at least one gene in Gs; (2) screening (M3)
 CC for an agent capable of modulating GCA or an inflammation (especially
 CC chronic) in a tissue, an allergic response in a subject, exposure of a
 CC subject to a pathogen or sterile inflammatory disease using the
 CC gene expression profile; (3) detecting (M4) an inflammation (especially
 CC chronic) in a tissue, an allergic response in a subject, exposure of a
 CC subject to a pathogen or sterile inflammatory disease, by detecting the
 CC level of expression in a sample of the tissue of gene(s) from Gs, where
 CC the level of expression of the gene is indicative of inflammation;
 CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,
 CC an allergic response in a subject, exposure of a subject to a pathogen
 CC or sterile inflammatory disease, by contacting a tissue having
 CC inflammation with an agent that modulates the expression of gene(s)
 CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
 CC modulating Gs; M3 is useful for screening an agent capable of modulating
 CC GCA preferably in an inflammation in a tissue; M4 is useful for
 CC detecting an inflammation (especially chronic) in a tissue, an allergic
 CC response in a subject, exposure of a subject to a pathogen or sterile
 CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
 CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
 CC reperfusion injury, ARDS, adult respiratory distress syndrome,
 CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,
 CC periodontal disease; also bacterial infection, viral infection,
 CC parasitic infection, protozoal infection, fungal infection and M5 is
 CC useful for treating one of the above conditions. The present
 CC sequence represents a gene differentially expressed in granulocytes.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 2691 BP; 600 A; 780 C; 831 G; 480 T; 0 other;

Query Match 100.0%; Score 215; DB 24; Length 2691;

Best Local Similarity 100.0%; Pred. No. 6.3e-55;

Matches 215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATACGGCTAGCTGAGGAGCTGCTGCGACAGTCCACTACCTTTTCGAGAGTGACTCC 60
 DB 274 ATACGGCTAGCTGAGGAGCTGCTGCGACAGTCCACTACCTTTTCGAGAGTGACTCC 333
 QY 61 GTTGTCCCAAGGCTTCCAGAGCGAACCTGTGCGGATCCAGTGTTCGAGTTCGAGTTT 120
 DB 334 GTTGTCCCAAGGCTTCCAGAGCGAACCTGTGCGGATCCAGTGTTCGAGTTCGAGTTT 393
 QY 121 CCGGCGTCCGGAAGGACCGAGCTTCTTCGCGGATCCAGTGTTCGAGTTCGAGTTT 180
 DB 394 CCGGCGTCCGGAAGGACCGAGCTTCTTCGCGGATCCAGTGTTCGAGTTCGAGTTT 453
 QY 181 TCTCAGAGCCGAGCGGACAGAGAGGAGGAAACCGC 215
 DB 454 TCTCAGAGCCGAGCGGACAGAGAGGAGGAAACCGC 488

RESULT 4

ABA99140

ID ABA99140 standard; cDNA; 2691 BP.

XX

AC ABA99140;

XX

DT 23-MAY-2002 (first entry)

XX Human hsp72 encoding sequence.

XX Human; hsp72; heat shock protein; cytostatic; antibacterial;

XX antiparasitic; MHC class I; ss.

XX Homo sapiens.

XX US6331388-B1.

XX 18-DEC-2001.

XX 17-OCT-1997; 97US-0955565.

XX 17-OCT-1997; 97US-0955565.

XX (WISC) WISCONSIN ALUMNI RES FOUND.

XX Malkovsky M, Wells AD;

XX WPI; 2002-138381/18.

XX Increasing expression of an MHC class I molecule in a cell, useful in

XX increasing antigen presentation and enhancing immune recognition of

XX cells infected with pathogens, by expressing a heat shock protein

XX introduced by a vector

XX Example 1; Fig 1; 89pp; English.

XX This invention relates to increasing expression of an MHC class I

XX molecule in a target cell, infected with a pathogen that is processed

XX by the MHC class I endogenous pathway. The method of expressing a

XX HSP is achieved by the introduction of an expression vector encoding

XX HSP to produce a transfected cell with increased expression of at least

XX one MHC class I molecule. The method is cytostatic, antibacterial

XX and antiparasitic. The method can be used to increase expression of an

XX MHC class I molecule in a target cell and to increase presentation of an

XX antigen on a cell surface by an MHC class I molecule. The method

XX can enhance the immunogenicity of the endogenous antigen in vivo, by

XX enhancing the generation of antibodies to an otherwise poorly

XX immunogenic antigen or cell. The method can be used in the

XX immunotherapy of cancer and pathogen infections, and for enhancing or

XX reducing radiation resistance of cells. The expression of the heat shock

XX protein in a cell enhances the presentation of endogenous antigens by

XX MHC class I molecules onto the cell surface in vitro, and can enhance

XX the immunogenicity of the endogenous antigen in vivo, by enhancing the

XX generation of antibodies to an otherwise poorly immunogenic antigen or

XX cell. This sequence represents the cDNA of human hsp72.

XX Sequence 2691 BP; 600 A; 780 C; 831 G; 480 T; 0 other;

Query Match 100.0%; Score 215; DB 24; Length 2691;

Best Local Similarity 100.0%; Pred. No. 6.3e-55;

Matches 215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATACGGCTAGCTGAGGAGCTGCTGCGACAGTCCACTACCTTTTCGAGAGTGACTCC 60
 DB 274 ATACGGCTAGCTGAGGAGCTGCTGCGACAGTCCACTACCTTTTCGAGAGTGACTCC 333
 QY 61 GTTGTCCCAAGGCTTCCAGAGCGAACCTGTGCGGATCCAGTGTTCGAGTTCGAGTTT 120
 DB 334 GTTGTCCCAAGGCTTCCAGAGCGAACCTGTGCGGATCCAGTGTTCGAGTTCGAGTTT 393
 QY 121 CCGGCGTCCGGAAGGACCGAGCTTCTTCGCGGATCCAGTGTTCGAGTTCGAGTTT 180
 DB 394 CCGGCGTCCGGAAGGACCGAGCTTCTTCGCGGATCCAGTGTTCGAGTTCGAGTTT 453
 QY 181 TCTCAGAGCCGAGCGGACAGAGAGGAGGAAACCGC 215
 DB 454 TCTCAGAGCCGAGCGGACAGAGAGGAGGAAACCGC 488

RESULT 5

AAT58086
 ID AAT58086 standard; cDNA; 2465 BP.
 XX
 AC AAT58086;
 XX
 DT 24-OCT-1997 (first entry)
 XX
 DE Human heat shock protein 70 cDNA.
 XX
 KW Human; heat shock protein 70; HSP70; primer; probe; detection;
 KW intracellular; abnormal transcription; acute; chronic; sustained;
 KW stress; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 263..2185
 FT /*tag= a
 FT /product= HSP70
 XX
 PN JP08322577-A.
 XX
 PD 10-DEC-1996.
 XX
 PF 01-JUN-1995; 95JP-0158581.
 XX
 PR 01-JUN-1995; 95JP-0158581.
 XX
 PA (HOKE-) HOKEN KAGAKU KENKYUSHO KK.
 XX
 DR WPI; 1997-081088/08.
 DR P-PSDB; AAW10065.
 XX
 PT Detection of abnormal transcription of HSP70 mRNA - using HSP70
 PT specific primer or probe, used in detection of human acute and
 PT chronic sustained stress load
 XX
 PS Claim 1; Fig 1; 13pp; Japanese.
 XX
 CC The present sequence, which encodes human heat shock protein 70
 CC (HSP70), is located on human chromosome 6p 21.3-22 and 14q 22-24
 CC and 21. Primers and probes based on the HSP70 cDNA coding
 CC sequence can be used to detect the abnormal transcription of
 CC intracellular HSP70 mRNA in human acute and chronic sustained
 CC stress load.
 XX
 SQ Sequence 2465 BP; 553 A; 713 C; 758 G; 441 T; 0 other;
 Query Match 99.3%; Score 213.4; DB 18; Length 2465;
 Best Local Similarity 99.5%; Pred. No. 1.9e-54;
 Matches 214; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 ATACGGCTAGCCTGAGAGCTGCTGCAGACAGTCCACTACCTTTTCGAGAGTGAATCCC 60
 DB 48 ATACGGCTAGCCTGAGAGCTGCTGCAGACAGTCCACTACCTTTTCGAGAGTGAATCCC 107
 QY 61 GTTGTCCCAAGCTTTCCAGAGCGAACTGTGCGGCTCGAGGACCGCGCGTGCAGTTT 120
 DB 108 GTTGTCCCAAGCTTTCCAGAGCGAACTGTGCGGCTCGAGGACCGCGCGTGCAGTTT 167
 QY 121 CCGGCGTCCGGAAGACCGAGCTCTTCGCGGATCCAGTTCCTGTTTCCAGCCCCCAA 180
 DB 168 CCGGCGTCCGGAAGACCGAGCTCTTCGCGGATCCAGTTCCTGTTTCCAGCCCCCAA 227
 QY 181 TCTCAGAGCCGAGCGCCAGACAGAGACAGGGAACCCG 215
 DB 228 TCTCAGAGCCGAGCGCCAGACAGAGACAGGGAACCCG 262
 RESULT 6
 AAA97541
 ID AAA97541 standard; cDNA; 2700 BP.

XX AAA97541;
 AC 29-JAN-2001 (first entry)
 DT Human Hsp72 (heat shock protein 72) cDNA.
 DE
 KW Human Hsp72; heat shock protein 72; chromosome 6p21.3; Hsp72 inhibitor;
 KW expression modulator; JNK phosphatase inhibitor; antiproliferative;
 KW drug screening; cancer; leukaemia; lymphoma; solid tumour; sarcoma;
 KW carcinoma; breast cancer; prostate cancer; premalignant condition; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200054814-A1.
 XX
 PD 21-SEP-2000.
 XX
 PF 17-MAR-2000; 2000WO-US07350.
 XX
 PR 18-MAR-1999; 99US-0125046.
 XX
 PA (PHYL-) PHYLOGENY INC.
 XX
 PI Volloch VZ, Sherman M;
 XX
 DR WPI; 2000-647056/62.
 DR P-PSDB; AAB23252.
 XX
 PT Identifying compounds that inhibit proliferation of cells and capable
 PT of modulating the expression of heat shock protein 72 gene and/or
 PT activity of Hsp72 useful for treating cancers such as leukemia,
 PT lymphoma
 XX
 PS Examples; Fig 16A; 77pp; English.
 XX
 CC The invention relates to a novel method of identifying compounds that
 CC inhibit proliferation of cells comprising contacting a test compound with
 CC a cell which overexpresses Hsp72 (heat shock protein 72), and determining
 CC if the test compound inhibits activity or expression of Hsp72.
 CC Optionally, Hsp72 is contacted with the test compound under optimum
 CC conditions to allow the two components to interact and bind, forming a
 CC complex which is detected. The invention also relates to a method of
 CC identifying compounds that inhibit Hsp72-mediated JNK phosphatase
 CC activation, comprising contacting a test compound with a cell which
 CC expresses Hsp72, exposing the cell to a heat induced stress and
 CC determining if the compound inhibits JNK phosphatase activity. The
 CC invention additionally encompasses compositions comprising an inhibitor
 CC of Hsp72 or JNK phosphatase activity. The compounds identified as
 CC inhibitors of Hsp72 or JNK phosphatase activity are useful for
 CC inhibiting the proliferation of cells. Modulation of the activity of the
 CC JNK phosphatase or Hsp72 is used to treat a proliferative disorder such
 CC as cancers (e.g., leukaemia, lymphoma, solid tumours such as sarcomas and
 CC carcinomas, breast cancer, prostate cancer). The compounds that inhibit
 CC Hsp72 activity can also be administered to treat premalignant conditions
 CC and to prevent progression to a neoplastic or malignant state. The
 CC compounds that inhibit Hsp72 function are administered to a patient
 CC having a disease or disorder mediated by an increase of Hsp72 expression
 CC or activity relative to normal levels. The present sequence represents
 CC cDNA encoding human Hsp72 used in the exemplifications of the invention.
 XX
 SQ Sequence 2700 BP; 601 A; 780 C; 834 G; 485 T; 0 other;
 Query Match 98.8%; Score 212.4; DB 21; Length 2700;
 Best Local Similarity 99.5%; Pred. No. 3.8e-54;
 Matches 213; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 ATACGGCTAGCCTGAGAGCTGCTGCAGACAGTCCACTACCTTTTCGAGAGTGAATCCC 60
 DB 274 ATACGGCTAGCCTGAGAGCTGCTGCAGACAGTCCACTACCTTTTCGAGAGTGAATCCC 333
 QY 61 GTTGTCCCAAGCTTTCCAGAGCGAACTGTGCGGCTGCAGGACCGCGCGTGCAGTTT 120

Db 334 GTTGTCCCAAGGCTTCCAGACGGAACCTGTGCGGTGTCAGGCACCGCGCGTCGAGTTT 393
QY 121 CCGGGTCCGGAAGACCGAGCTCTTCTCGCGGATCCAGTGTTCGGTTTCAGCCCCCAA 180
Db 394 CCGGGTCCGGAAGACCGAGCTCTTCTCGCGGATCCAGTGTTCGGTTTCAGCCCCCAA 453
QY 181 TCTCAGACCGGACCGACGACGAGGAGGAAACCG 214
Db 454 TCTCAGACCGGACCGACGACGAGGAGGAAACCG 487

RESULT 7
AAAI5620
ID AAAI5620 standard; DNA; 2465 BP.
XX AC AAAI5620;
XX DT 31-JUL-2000 (first entry)
XX DE Human heat shock protein HSP70 nucleotide sequence.
XX KW Human; heat shock protein; HSP70; chromosome 6p21.3-22; stress;
KW chromosome 14q22-24; transcription; rheumatism; schizophrenia;
KW depression; nephrotic syndrome; ss.
XX OS Homo sapiens.
XX PN JP2000069999-A.
XX PD 07-MAR-2000.
XX PF 01-JUN-1995; 99JP-0257146.
XX PR 01-JUN-1995; 95JP-0158581.
XX XX (HOKE-) HOKEN KAGAKU KENKYUSHO KK.
XX WPI; 2000-264458/23.
XX DR P-PSDB; AAY88408.
XX XX
XX PT Abnormal transcription of intracellular HSP70mRNA under acute and
PT chronic continuous load of stress in a human being and its application
PT
XX PS Claim 2; Fig 1; lipp; Japanese.
XX CC This sequence represents the human heat shock protein HSP70 nucleotide
CC sequence. Human heat shock proteins are located on chromosomes 6p21.3-22
CC and 14q22-24. The invention relates to the abnormal transcription of
CC intracellular HSP70mRNA under acute and chronic stress load in a human.
CC The abnormal transcription of HSP70 can be used in the improvement of
CC stress and response and diagnosis of stress diseases including
CC rheumatism, schizophrenia, depression and nephrotic syndrome.
XX SQ Sequence 2465 BP; 553 A; 722 C; 749 G; 441 T; 0 other;
Query Match 97.0%; Score 208.6; DB 21; Length 2465;
Best Local Similarity 98.1%; Pred. No. 5.2e-53;
Matches 211; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATACGGCTAGCTGAGGAGCTGCTGCGAGTCCACTACCTTTTCGAGAGTGACTCCC 60
Db 48 ATACGGCTAGCTGAGGAGCTGCTGCGAGTCCACTACCTTTTCGAGAGTGACTCCC 107
QY 61 GTTGTCCCAAGGCTTCCAGAGGAACTGTGCGGTGAGGACCGGCGCTCGAGTTT 120
Db 108 GTTGTCCCAAGGCTTCCAGAGGAACTGTGCGGTGAGGACCGGCGCTCGAGTTT 167
QY 121 CCGGGTCCGGAAGGACCGAGCTCTTCTCGCGGATCCAGTGTTCGGTTTCAGCCCCCAA 180
Db 168 CCGGGTCCGGAAGGACCGAGCTCTTCTCGCGGATCCAGTGTTCGGTTTCAGCCCCCAA 227
QY 181 TCTCAGACCGGACCGACGACGAGGAGGAAACCG 215

Db 228 TCTCAGACCGGACCGACGAGGAGGAAACCG 262

RESULT 8
AAAI5622
ID AAAI5622 standard; DNA; 2302 BP.
XX AC AAAI5622;
XX DT 31-JUL-2000 (first entry)
XX DE Human heat shock protein SHSP70 nucleotide sequence.
XX KW Human; heat shock protein; HSP70; chromosome 6p21.3-22; stress;
KW chromosome 14q22-24; transcription; rheumatism; schizophrenia;
KW depression; nephrotic syndrome; SHSP70; ss.
XX OS Homo sapiens.
XX PN JP2000069999-A.
XX PD 07-MAR-2000.
XX PF 01-JUN-1995; 99JP-0257146.
XX PR 01-JUN-1995; 95JP-0158581.
XX XX (HOKE-) HOKEN KAGAKU KENKYUSHO KK.
XX WPI; 2000-264458/23.
XX DR P-PSDB; AAY88410.
XX XX
XX PT Abnormal transcription of intracellular HSP70mRNA under acute and
PT chronic continuous load of stress in a human being and its application
PT
XX PS Disclosure; Fig 3; lipp; Japanese.
XX CC This sequence represents the human heat shock protein SHSP70 nucleotide
CC sequence. Human heat shock proteins are located on chromosomes 6p21.3-22
CC and 14q22-24. The invention relates to the abnormal transcription of
CC intracellular HSP70mRNA under acute and chronic stress load in a human.
CC The abnormal transcription of HSP70 can be used in the improvement of
CC stress and response and diagnosis of stress diseases including
CC rheumatism, schizophrenia, depression and nephrotic syndrome.
XX SQ Sequence 2302 BP; 515 A; 652 C; 712 G; 423 T; 0 other;
Query Match 86.9%; Score 186.8; DB 21; Length 2302;
Best Local Similarity 92.1%; Pred. No. 1.9e-46;
Matches 197; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 1 ATACGGCTAGCTGAGGAGCTGCTGCGAGTCCACTACCTTTTCGAGAGTGACTCCC 60
Db 48 ATACGGCTAGCTGAGGAGCTGCTGCGAGTCCACTACCTTTTCGAGAGTGACTCCC 107
QY 61 GTTGTCCCAAGGCTTCCAGAGGAACTGTGCGGTGAGGACCGGCGCTCGAGTTT 120
Db 108 GTTGTCCCAAGGCTTCCAGAGGAACTGTGCGGTGAGGACCGGCGCTCGAGTTT 167
QY 121 CCGGGTCCGGAAGGACCGAGCTCTTCTCGCGGATCCAGTGTTCGGTTTCAGCCCCCAA 180
Db 168 CCGGGTCCGGAAGGACCGAGCTCTTCTCGCGGATCCAGTGTTCGGTTTCAGCCCCCAA 227
QY 181 TCTCAGACCGGACCGACGACGAGGAGGAAACCG 214
Db 228 TCTCAGACCGGACCGAGCGGCTCATCGGGGATCCG 261

RESULT 9
AAH8560/c
ID AAH8560 standard; DNA; 511 BP.


```
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX Homo sapiens.
OS
XX WO200218632-A2.
XX 07-MAR-2002.
XX 01-SEP-2001; 2001WO-EP10074.
XX 01-SEP-2000; 2000DE-1043826.
XX 05-SEP-2000; 2000DE-1044543.
XX (EPIG-) EPIGENOMICS AG.
XX Olek A, Piepenbrock C, Berlin K, Guetig D;
PI WPI; 2002-371829/40.
XX Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA
XX Claim 12; 56pp + Sequence Listing; 56pp; German.
XX This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX Sequence 595 BP; 91 A; 56 C; 186 G; 262 T; 0 other;
XX Query Match 67.5%; Score 145.2; DB 24; Length 595;
XX Best Local Similarity 79.9%; Pred. No. 4.6e-34;
XX Matches 171; Conservative 0; Mismatches 43; Indels 0; Gaps 0;
XX 1 ATAACGGCTAGCTGAGGAGCTGCTGCAGACGCTCCACTACCTTTTCGAGAGTGACTCCC 60
XX 454 ATAACGACTTAACCTTAAACAACTACTACGACAACTCACTACCTTTTCGAAATAACTCCC 395
XX 61 GTTGTCCCAAGGCTTCCAGAGCGAACCTGTGCGGCTGCAGGACCGCGCGTCCGAGTTT 120
XX 394 GTTATCCCAAACTTCCCAAAACGAACTTATACGACTACAAACCGGCGCTCGAATTT 335
XX 121 CCGGCGTCCGGAAGGACCGAGCTTTCTCGCGGATCCAGTTCCTGTTTCCAGGCCCA 180
XX 334 CCGAGCTCCGGAAGGACCGAGCTTTCTCGCGAATCCAAATATTCGTTTCCAAACCCCA 275
XX 181 TCTCAGAGCCGAGCGGACAGAGGAGGAAACCG 214
XX 274 TCTCAAAACCGAACCGACAAAACAAAACCG 241
```

```
RESULT 12
ABQ30839
ID ABQ30839 standard; DNA; 595 BP.
XX AC ABQ30839;
XX 12-JUL-2002 (first entry)
XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 17430.
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX Homo sapiens.
XX WO200218632-A2.
XX 07-MAR-2002.
XX 01-SEP-2001; 2001WO-EP10074.
XX 01-SEP-2000; 2000DE-1043826.
XX 05-SEP-2000; 2000DE-1044543.
XX (EPIG-) EPIGENOMICS AG.
XX Olek A, Piepenbrock C, Berlin K, Guetig D;
PI WPI; 2002-371829/40.
XX Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA
XX Claim 12; 56pp + Sequence Listing; 56pp; German.
XX This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX Sequence 595 BP; 262 A; 186 C; 56 G; 91 T; 0 other;
XX Query Match 67.5%; Score 145.2; DB 24; Length 595;
XX Best Local Similarity 79.9%; Pred. No. 4.6e-34;
XX Matches 171; Conservative 0; Mismatches 43; Indels 0; Gaps 0;
XX 1 ATAACGGCTAGCTGAGGAGCTGCTGCAGACGCTCCACTACCTTTTCGAGAGTGACTCCC 60
XX 142 ATAACGACTTAACCTTAAACAACTACTACGACAACTCACTACCTTTTCGAAATAACTCCC 201
XX 61 GTTGTCCCAAGGCTTCCAGAGCGAACCTGTGCGGCTGCAGGACCGCGCGTCCGAGTTT 120
XX 202 GTTATCCCAAACTTCCCAAAACGAACTTATACGACTACAAACCGGCGCTCGAATTT 261
```


CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.

XX
SQ Sequence 596 BP; 264 A; 186 C; 55 G; 91 T; 0 other;

Query Match 67.5%; Score 145.2; DB 24; Length 596;
Best Local Similarity 79.9%; Pred. No. 4.6e-34;
Matches 171; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 1 ATAACGGCTAGCTCGAGGAGCTGCTGCACAGTCCACTACCTTTTCGAGAGTGACTCCC 60
DB 52 ATAACGACTAACTTAAATAAATACTACGACAAATCCACTACCTTTTCGAAATAAATACTCCC 111
QY 61 GTTGTCCCAAGGCTTCCAGAGCGAACTGTGCGGCTGCAGGCACCGCGCGTTCGAGTTT 120
DB 112 GTTATCCCAAAACTTCCCAAAAGAACTATAGACTACAACACCGAGCGTTCGAATTT 171
QY 121 CCGGCGTCCGGAAGACGAGCTTCTTCGCGGATCCAGTGTTCGTTTCCAGCCCCCAA 180
DB 172 CCGAGCTCCGAAAAAACCAGAACTCTTCTCGCGAATCCAATATTCGGTTTCCAAACCCCAA 231
QY 181 TCTCAGACCGAGCCGACAGAGCAGGGAACCG 214
DB 232 TCTCAAAACCGAACCGACAAAAAACAAAAAACCG 265

RESULT 15

ABL34442/C
ID ABL34442 standard; DNA; 9899 BP.

XX ABL34442;

XX 26-MAR-2002 (first entry)

XX Human immune system associated gene SEQ ID NO: 2415.

XX Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.

XX Homo sapiens.

XX WO200200928-A2.

XX 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP07537.

XX 30-JUN-2000; 2000DE-1032529.

XX 01-SEP-2000; 2000DE-1043826.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -

XX Claim 1; SEQ ID NO 2415; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and

CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.

XX
SQ Sequence 9899 BP; 2386 A; 359 C; 2673 G; 4481 T; 0 other;

Query Match 67.5%; Score 145.2; DB 24; Length 9899;
Best Local Similarity 79.9%; Pred. No. 8.5e-34;
Matches 171; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 1 ATAACGGCTAGCTCGAGGAGCTGCTGCACAGTCCACTACCTTTTCGAGAGTGACTCCC 60
DB 3406 ATAACGACTAACTTAAATAAATACTACGACAAATCCACTACCTTTTCGAAATAAATACTCCC 3347
QY 61 GTTGTCCCAAGGCTTCCAGAGCGAACTGTGCGGCTGCAGGCACCGCGGTCGAGTTT 120
DB 3346 GTTATCCCAAACTTCCCAAAAGAACTATAGACTACAACACCGAGCGTTCGAATTT 3287
QY 121 CCGGCGTCCGGAAGACGAGCTCTTCTCGCGGATCCAGTGTTCGTTTCCAGCCCCCAA 180
DB 3286 CCGAGCTCCGAAAAAACCAGAACTCTTCTCGCGAATCCAATATTCGGTTTCCAAACCCCAA 3227
QY 181 TCTCAGACCGAGCCGACAGAGCAGGGAACCG 214
DB 3226 TCTCAAAACCGAACCGACAAAAAACAAAAAACCG 3193

Search completed: February 3, 2004, 02:38:49
Job time : 178 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: /February 3, 2004, 03:00:42 ; Search time 1635 Seconds
(without alignments)
484.392 Million cell updates/sec

Title: US-09-936-506-1
Perfect score: 215
Sequence: 1 ataacggttagcctgaggag.....sacagagacgaggaaccgc 215

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 2449703 seqs, 1841816367 residues

Total number of hits satisfying chosen parameters: 4899406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubnpa/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubnpa/PTC_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubnpa/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubnpa/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubnpa/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubnpa/PTC_NEW_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubnpa/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubnpa/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubnpa/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubnpa/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubnpa/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/2/pubnpa/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/2/pubnpa/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/2/pubnpa/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubnpa/US10_NEW_PUB.seq:*
- 16: /cgn2_6/ptodata/2/pubnpa/US60_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/2/pubnpa/US60_NEW_PUBCOMB.seq:*
- 18: /cgn2_6/ptodata/2/pubnpa/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	213.4	99.3	549	12	US-10-348-359-1
2	210.4	97.9	1903	12	US-10-108-260A-1023
3	191.4	89.0	2412	11	US-09-919-039-144
4	163.6	76.1	506	13	US-10-029-386-6557
5	156.8	72.9	2458	11	US-09-919-039-145
6	156.4	72.7	2771	13	US-10-027-632-111980
7	156.4	72.7	2771	14	US-10-027-632-111980
8	145.2	67.5	9899	13	US-10-311-455-2415
9	145.2	67.5	3673778	13	US-10-312-841-2
10	134	62.3	9899	13	US-10-311-455-2416
11	134	62.3	3673778	13	US-10-312-841-1
12	58	27.0	939	13	US-10-029-386-20724
13	38.4	17.9	2636	14	US-10-002-600-95
14	34.2	15.9	2940917	13	US-10-027-632-174763
15	34.2	15.9	2940917	14	US-10-027-632-174763

16	32.2	15.0	9698	12	US-10-210-130-37	Sequence 37, Appl
17	32.2	15.0	9807	12	US-10-379-381-1	Sequence 1, Appl
18	32.2	15.0	9930	12	US-10-210-130-33	Sequence 33, Appl
19	32.2	15.0	10122	12	US-10-210-130-35	Sequence 35, Appl
20	32.2	15.0	62805	12	US-10-379-381-3	Sequence 3, Appl
21	32	14.9	3264	13	US-09-894-159-5	Sequence 5, Appl
22	31.8	14.8	4722	9	US-09-962-055-14	Sequence 14, Appl
23	31.8	14.8	4722	10	US-09-976-740-14	Sequence 14, Appl
24	31.8	14.8	4722	12	US-10-616-187-14	Sequence 14, Appl
25	31.8	14.8	4722	14	US-10-023-529-14	Sequence 14, Appl
26	31.8	14.8	4722	14	US-10-023-523-14	Sequence 14, Appl
27	31.6	14.7	391	10	US-09-880-107-2763	Sequence 2763, Ap
28	31	14.4	9025608	15	US-10-156-761-1	Sequence 1, Appl
29	30.6	14.2	381	12	US-10-242-535A-26318	Sequence 26318, A
30	30.6	14.2	9025608	15	US-10-156-761-1	Sequence 1, Appl
31	30.4	14.1	981	15	US-10-156-761-3761	Sequence 3761, Ap
32	30.2	14.0	439	9	US-09-772-105-50	Sequence 50, Appl
33	30.2	14.0	888	15	US-10-156-761-2771	Sequence 2771, Ap
34	30.2	14.0	3086	12	US-10-292-798-609	Sequence 609, App
35	30.2	14.0	3086	13	US-10-017-161-697	Sequence 697, App
36	30	14.0	1589	12	US-10-108-260A-2104	Sequence 2104, Ap
37	30	14.0	1746	15	US-10-156-761-4695	Sequence 4695, Ap
38	30	14.0	10419	13	US-10-029-120-3	Sequence 3, Appl
39	30	14.0	10419	14	US-10-027-806-3	Sequence 3, Appl
40	30	14.0	10419	14	US-10-034-623-3	Sequence 3, Appl
41	30	14.0	10419	15	US-10-027-801-3	Sequence 3, Appl
42	30	14.0	42432	13	US-10-029-120-2	Sequence 2, Appl
43	30	14.0	42432	14	US-10-027-806-2	Sequence 2, Appl
44	30	14.0	42432	14	US-10-034-623-2	Sequence 2, Appl
45	30	14.0	42432	15	US-10-027-801-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1

US-10-348-359-1
Sequence 1, Application US/10348359
Publication No. US20040018178A1
GENERAL INFORMATION:
APPLICANT: WEST, MICHAEL
TITLE OF INVENTION: STEM CELL-DERIVED ENDOTHELIAL CELLS MODIFIED TO DISRUPT
FILE REFERENCE: 100375.54374US
CURRENT APPLICATION NUMBER: US/10/348,359
CURRENT FILING DATE: 2003-01-22
PRIOR APPLICATION NUMBER: 60/349,345
PRIOR FILING DATE: 2002-01-22
NUMBER OF SEQ ID NOS: 6
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 549
TYPE: DNA
ORGANISM: Homo sapiens
US-10-348-359-1

Query Match	99.3%	Score	213.4	DB 12	Length	549			
Best Local Similarity	99.5%	Pred. No.	3.1e-59						
Matches	214	Conservative	0	Mismatches	1	Indels	0	Gaps	0
QY	1	ATAACGGCTAGCCTGAGGAGCTGCTCGACAGTCCACTACCTTTTTCGAGAGTGAATCCC	60						
Db	269	ATAACGGCTAGCCTGAGGAGCTGCTCGACAGTCCACTACCTTTTTCGAGAGTGAATCCC	328						
QY	61	GTGTGTCCTCAAGGCTTCCCGAGAGCGAACCTGTGCGGCTGCAGGCACCGCGCGTCTGAGTTT	120						
Db	329	GTGTGTCCTCAAGCTTCCCGAGAGCGAACCTGTGCGGCTGCAGGCACCGCGCGTCTGAGTTT	388						
QY	121	CGGCGCTCCGGAGAGCCGAGCTCTTCTCGCGGATCCAGTGTTCGTTTTCGAGCCCCCAA	180						
Db	389	CGGCGCTCCGGAGAGCCGAGCTCTTCTCGCGGATCCAGTGTTCGTTTTCGAGCCCCCAA	448						
QY	181	TCTCAGAGCCGAGCCGACAGAGAGCGGAACCGC	215						

```
Db 449 TCTCAGAGCGGAGCGGACAGAGAGCAGGGAACCG 483
|||||
RESULT 2
US-10-108-260A-1023
; Sequence 1023, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20040005560A1el full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1023
; LENGTH: 1903
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-108-260A-1023

Query Match 97.9%; Score 210.4; DB 12; Length 1903;
Best Local Similarity 99.5%; Pred. No. 3.1e-58;
Matches 211; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AACGCTAGCCTGAGAGCTGCTGGAGAGTCCACTACTTTTTCGAGAGTGACTCCCGT 62
|||||
Db 1 AACGCTAGCCTGAGAGCTGCTGGAGAGTCCACTACTTTTTCGAGAGTGACTCCCGT 60
|||||
QY 63 TGTCCTCAAGGCTTCCAGAGCGAACTGTGCGGCTGCAGGACCGCGCGCTCGAGTTTCC 122
|||||
Db 61 TGTCCTCAAGGCTTCCAGAGCGAACTGTGCGGCTGCAGGACCGCGCGCTCGAGTTTCC 120
|||||
QY 123 GCGCTCCGAGAGCGAGCTTCTCGCGATCCAGTGTCCGTTTCAGCCCCCAATC 182
|||||
Db 121 GCGCTCCGAGAGCGAGCTTCTCGCGATCCAGTGTCCGTTTCAGCCCCCAATC 180
|||||
QY 183 TCAGAGCCGAGCCGACAGAGCAGGGAACCG 214
|||||
Db 181 TCAGAGCCGAGCCGACAGAGCAGGGAACCG 212
|||||

RESULT 3
US-09-919-039-144
; Sequence 144, Application US/09919039
; Publication No. US20030108871A1
; GENERAL INFORMATION:
; APPLICANT: Kaser, Matthew R.
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
; FILE REFERENCE: PA-0035 US
; CURRENT APPLICATION NUMBER: US/09/919,039
; CURRENT FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 60/222,113
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 401
; SOFTWARE: PERL Program
; SEQ ID NO 144
; LENGTH: 2412
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030108871A1 242010.16
US-09-919-039-144

Query Match 89.0%; Score 191.4; DB 11; Length 2412;
Best Local Similarity 99.0%; Pred. No. 4.4e-52;
Matches 203; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 11 GCGTGAAGAGCTGCTGCGACAGTCCACTACTTTTTCGAGAGTGACTCCCGTTGTCCAA 70
|||||
Db 1 GCGTGAAGAGCTGCTGCGACAGTCCACTACTTTTTCGAGAGTGACTCCCGTTGTCCAA 60
|||||

QY 71 GGCTTCCAGAGCGAACTGT-GCGGCTGCAGGACCGGCGCGTTCGAGTTTCCGCGCTCC 129
|||||
Db 61 GGCTTCCAGAGCGAACTGTGGCGGCTGCAGGACCGGCGCGTTCGAGTTTCCGCGCTCC 120
|||||
QY 130 GGAAGGACCGAGCTCTTCTCGCGATCCAGTGTTCGTTTCCAGCCCCCAATCTCAGAGC 189
|||||
Db 121 GGAAGGACCGAGCTCTTCTCGCGATCCAGTGTTCGTTTCCAGCCCCCAATCTCAGAGC 180
|||||
QY 190 CGAGCCGACAGAGCAGGGAACCG 214
|||||
Db 181 GGAGCCGACAGAGCAGGGAACCG 205
|||||

RESULT 4
US-10-029-386-6557/c
; Sequence 6557, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David K.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; FILE REFERENCE: AEMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 6557
; LENGTH: 506
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AF134726.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 5.5
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.8
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 11
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 7.9
; OTHER INFORMATION: SWISSPROT HIT: P08107, EVALUATE 2.00e-17
; OTHER INFORMATION: NT HIT: M59830.1, EVALUATE 0.00e+00
; OTHER INFORMATION: EST_HUMAN HIT: BG773197.1, EVALUATE 0.00e+00
US-10-029-386-6557

Query Match 76.1%; Score 163.6; DB 13; Length 506;
Best Local Similarity 88.1%; Pred. No. 3.8e-43;
Matches 178; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 1 ATAAACGGCTAGCCTGAGGAGCTGCTGCGACAGTCCACTACTTTTTCGAGAGTGACTCC 60
|||||
Db 339 AAAACGGCCAGCCTGAGGAGCTGCTGCGAGGCTCGCTTCGAGAGTGACTCC 280
|||||
QY 61 GTTGTCCCAAGGCTTCCAGAGCGAACTGTGCGGCTGCAGGACCGCGCGTTCGAGTTT 120
|||||
Db 279 GCGGTCCCAAGGCTTTCAGAGCGAACTGTGCGGCTGCAGGACCGCGGTTGAGTTT 220
|||||
QY 121 CCGGCTCCGGAAGGACCGAGCTCTTCTCGCGGATCCAGTGTTCGTTTCCAGCCCCCAA 180
|||||
Db 219 CCGGCTTCCGAAGGAGCTGAGCTCTTGTGCGGATCCCGCTCCGCGGTTTCAGCCCCCAG 160
|||||
QY 181 TCTCAGAGCCGAGCCGACAGAG 202
|||||
Db 159 TCTCAGAGCCGAGCCGACAGAG 138
|||||

RESULT 5
US-09-919-039-145
; Sequence 145, Application US/09919039
; Publication No. US20030108871A1
; GENERAL INFORMATION:
; APPLICANT: Kaser, Matthew R.
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
```



```

; FILE REFERENCE: PA-0035 US
; CURRENT APPLICATION NUMBER: US/09/919,039
; CURRENT FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 60/222,113
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 401
; SOFTWARE: PERL Program
; SEQ ID NO 145
; LENGTH: 2458
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20030108871A1 1678695CBI
; US-09-919-039-145

Query Match          72.9%; Score 156.8; DB 11; Length 2458;
Best Local Similarity 88.5%; Pred. No. 6.6e-41;
Matches 170; Conservative 0; Mismatches 22; Indels 0; Gaps 0

Qy 11 GCCTGAGGAGCTGCTGCACAGTCCACTACCTTTTTCGAGAGTGACTCCCGTTGTCCCAA 70
Db 1 GCCTGAGGAGCTGCTGCAGGGTCCGCTTGGTCTTCGAGAGTGACTCCCGCGGTCCCA 60

Qy 71 GCGTTCACAGACGAACTGTGCGGCTCGAGGACACGGCGCGCTCGAGTTTCGGGCGTCCG 130
Db 61 GCGTTTTCAGAGCGAACTGTGCGGCTCGAGGACACGGCGCGTGTGAGTTTCGGGCGTCC 120

Qy 131 GAAGGACCGAGCTCTTCGCGGATCCAGTGTTCGGTTTCAGAGCCCCCAATCTCAGAGCC 190
Db 121 GAAGGACTGAGCTCTTGTGCGGATCCCGTCCGCGGTTTCAGAGCCCCCAGTCTCAGAGCG 180

Qy 191 GAGCCGACAGAG 202
Db 181 GAGCCGACAGAG 192

RESULT 6
US-10-027-632-111980/c
; Sequence 111980, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827,129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 111980
; LENGTH: 2771
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-111980

Query Match          72.7%; Score 156.4; DB 13; Length 2771;
Best Local Similarity 99.4%; Pred. No. 9e-41;
Matches 157; Conservative 0; Mismatches 1; Indels 0; Gaps 0

```



```
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; APPLICANT: Rank, David R.
; NAME/KEY: unsure
; LOCATION: (3294164)
US-10-312-841-1

Query Match      62.3%; Score 134; DB 13; Length 3673778;
Best Local Similarity 76.6%; Pred. No. 2.6e-33;
Matches 164; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 1 ATAAAGGCTAGCTGAGGAGCTGCTCCAGAGCAACCTGCGGCTGCAGGCAGCCGCGCTCGAGTTT 120
DB 2101291 ATAAAGGTTAGTTGAGGAGTTGTTGCGATAGTTTATTATTATTTTTCGAGAGTGATTTTC 2101350

QY 61 GTTGTCCCAAGGCTTCCAGAGCAACCTGCGGCTGCAGGCAGCCGCGCTCGAGTTT 120
DB 2101351 GTTGTTTAAGTTTTCAGAGCAATTTGTCGGTTGTAGTATCGGCGCGCTCGAGTTT 2101410

QY 121 CCGGCTCCGGAAGCAACGAGCTCTTCTCGCGATCCAGTGTTCGGTTTCCAGCCCCCAA 180
DB 2101411 TCGGCGTTCGGAAGCATCGATTTTTCGCGATTTAGTGTTCGTTTATGTTTAA 2101470

QY 181 TCTCAGCGCGAGCCGACAGAGAGGGAACCG 214
DB 2101471 TTTTAGAGTCGAGTCGATAGAGTAGGGAATCG 2101504

RESULT 12
US-10-029-386-20724/c
; Sequence 20724, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Rank, David R.
; APPLICANT: Penn, Sharron G.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEONICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 20724
; LENGTH: 939
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AF134726.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 11
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 9.6
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.6
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 31
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1e+02
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 17
; OTHER INFORMATION: NT HIT: g116174860, EVALUE 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: P08107, EVALUE 0.00e+00
; OTHER INFORMATION: EST_HUMAN HIT: BG697380.1, EVALUE 0.00e+00
US-10-029-386-20724

Query Match      27.0%; Score 58; DB 13; Length 939;
Best Local Similarity 100.0%; Pred. No. 5.2e-09;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 157 CAGTGTTCGGTTTCAGGCCCCCAATCTCAGAGCCGAGCCGACAGAGAGGGAACCG 214
DB 939 CAGTGTTCGGTTTCAGGCCCCCAATCTCAGAGCCGAGCCGACAGAGAGGGAACCG 882

RESULT 13
US-10-002-600-95
; Sequence 95, Application US/10002600
; Publication No. US20020137077A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Hopkins, Christopher M.
; APPLICANT: Peterson, David P.
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Hawkins, Phillip R.
; TITLE OF INVENTION: GENES REGULATED IN ACTIVATED T CELLS
; FILE REFERENCE: PA-0042 US
; CURRENT APPLICATION NUMBER: US/10/002,600
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: 60/243,521
; PRIOR FILING DATE: 2000-10-25
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: PERL Program
; SEQ ID NO 95
; LENGTH: 2636
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Template ID: 242010.34
US-10-002-600-95

Query Match      17.9%; Score 38.4; DB 14; Length 2636;
Best Local Similarity 97.5%; Pred. No. 0.012;
Matches 39; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 175 CCCCAATCTCAGAGCCGAGCCGACAGAGAGGGAACCG 214
DB 390 CCCCAATCTCAGAGCCGAGCCGACAGAGAGGGAACCG 429

RESULT 14
US-10-027-632-174763
; Sequence 174763, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 174763
; LENGTH: 2940917
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(2940917)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-174763

Query Match      15.9%; Score 34.2; DB 13; Length 2940917;
Best Local Similarity 53.5%; Pred. No. 0.45;
Matches 69; Conservative 1; Mismatches 59; Indels 0; Gaps 0;

QY 65 TCCCAAGGCTTCCAGAGCGAACTGTGCGGTGTCAGGACCGGCGGTTCGAGTTTCCGG 124
```

Db 2433019 TCCTAGTGGTCCCAAGATGTTGGTAGTCTGCTGCAATTCGGCTTGGCGGACCAAGTT 2433078
QY 125 CGTCCGGAAGACCGAGCTCTTCTCGCGGATCCAGTGTTCGGTTCCAGCCCCCAATCTC 184
Db 2433079 CCATCGTAGGACCGAGGACCTCTAGTGGTTCCAGGGTGGRTCTGCAAACTTAACCAAC 2433138
QY 185 AGAGCCGAG 193
Db 2433139 AGTCCAGAG 2433147

RESULT 15
US-10-027-632-174763
; Sequence 174763, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 174763
; LENGTH: 2940917
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(2940917)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-174763

Query Match 15.9%; Score 34.2; DB 14; Length 2940917;
Best Local Similarity 53.5%; Pred No. 0.45; Mismatches 59; Indels 0; Gaps 0;
Matches 69; Conservative 1;
QY 65 TCCCAAGGCTTCCAGAGCGAACCTGTGCGGCTGCAGGACCGGCGGTTCGAGTTTCCGG 124
Db 2433019 TCCTAGTGGTCCCAAGATGTTGGTAGTCTGCTGCAATTCGGCTTGGCGGACCAAGTT 2433078
QY 125 CGTCCGGAAGACCGAGCTCTTCTCGCGGATCCAGTGTTCGGTTCCAGCCCCCAATCTC 184
Db 2433079 CCATCGTAGGACCGAGGACCTCTAGTGGTTCCAGGGTGGRTCTGCAAACTTAACCAAC 2433138
QY 185 AGAGCCGAG 193
Db 2433139 AGTCCAGAG 2433147

Search completed: February 3, 2004, 04:45:24
Job time : 1662 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2004, 01:38:21 ; Search time 1226 Seconds

(without alignments)
4262.204 Million cell updates/sec

Title: US-09-936-506-1

Perfect score: 215

Sequence: 1 ataacgctagcctgaggag.....gacagagcaggaacgcg 215

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_est1:*

10: gb_est2:*

11: gb_est3:*

12: gb_est4:*

13: gb_est5:*

14: gb_est6:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_phg:*

27: em_gss_vrl:*

28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	212.4	98.8	356	14	CB112917 K-EST0154
2	212.4	98.8	493	13	B0639435 hel5ell.y
3	212.4	98.8	511	14	CB154948 K-EST0213
4	212.4	98.8	527	14	CB141518 K-EST0195

5	212.4	98.8	590	14	CB138426
6	212.4	98.8	701	12	BG773197
7	212.4	98.8	816	10	BG485554
8	212.4	98.8	916	12	BI462679
9	211.4	98.3	751	12	BI464037
10	210.4	97.9	267	9	AU076824
11	210.4	97.9	316	12	BM822968
12	210.4	97.9	376	14	CB133590
13	210.4	97.9	409	14	CB133685
14	210.4	97.9	436	12	BI548281
15	210.4	97.9	463	14	CB137170
16	210.4	97.9	479	14	CB133251
17	210.4	97.9	505	14	CB133452
18	210.4	97.9	509	14	CB134000
19	210.4	97.9	509	14	CB138961
20	210.4	97.9	516	14	CB137829
21	210.4	97.9	516	14	CB138598
22	210.4	97.9	522	14	CB133993
23	210.4	97.9	523	14	CB152005
24	210.4	97.9	523	14	CB152693
25	210.4	97.9	532	14	CB153107
26	210.4	97.9	533	14	CB133518
27	210.4	97.9	546	14	CB155190
28	210.4	97.9	588	12	BM849223
29	210.4	97.9	588	14	CB155167
30	210.4	97.9	588	14	CB156537
31	210.4	97.9	589	14	CB139057
32	210.4	97.9	606	10	BG719904
33	210.4	97.9	709	10	BG699144
34	210.4	97.9	772	10	BG703851
35	210.4	97.9	888	12	BI552752
36	210.4	97.9	904	12	BI552752
37	209.2	97.3	812	12	BI603727
38	208.8	97.1	498	12	BI550745
39	208.8	97.1	517	14	CB139964
40	208.8	97.1	768	12	BI544988
41	208.8	97.1	781	10	BG715565
42	208.8	97.1	896	12	BI600683
43	208.8	97.1	904	12	BI597081
44	208.8	97.1	916	12	BI561591
45	207.2	96.4	509	10	BG707724

ALIGNMENTS

RESULT 1
CB112917

LOCUS

DEFINITION K-EST0154803 L6ChoCKO Homo sapiens CDNA clone L6ChoCKO-10-All 5', linear EST 28-JAN-2003

ACCESSION CB112917

VERSION CB112917.1 GI:27938724

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

REFERENCE 1 (bases 1 to 356)

AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,Y.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R., Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and Kim,Y.S.

TITLE 21C Frontier Korean EST Project 2001

JOURNAL Unpublished

COMMENT Contact: Kim YS

Genome Research Center

Korea Research Institute of Bioscience & Biotechnology

52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea

Tel: +82-42-860-4470

Fax: +82-42-860-4409

Email: yongsung@mail.kribb.re.kr

Plate: 10 row: A column: 11

High quality sequence stop: 356.

FEATURES
source

Location/Qualifiers
1. .356
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="L6choCK0-10-A11"
/sex="M"
/cell_line="Cho-CK"
/lab_host="Top10P"
/clone_lib="L6choCK0"
/note="Organ: Liver; Vector: pCNS-D2; Site 1: EcoRI;
Site 2: NotI; The poly (A)+ RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then decapped
with tobacco acid pyrophosphatase (TAP). The decapped
intact mRNA was ligated with DNA-RNA linker including
EcoRI site by treatment of T4 RNA ligase and the first
strand cDNA was synthesized from oligo dt-selected mRNA by
priming with dt-tailed vector. The dt-tailed vector was
adjusted to have about 60nt. The cDNA vector was
circularized with E. coli DNA ligase after digestion of
EcoRI which site is also included in vector. An RNA strand
converted to a DNA strand by Okayama-Berg method. The
obtained cDNA vectors were used for transformation of
competent cells E. coli 'Top10P' by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library."

BASE COUNT
ORIGIN

71 a 121 c 105 g 59 t
Query Match 98.8%; Score 212.4; DB 14; Length 356;
Best Local Similarity 99.5%; Pred. No. 6.7e-46;
Matches 213; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATACGGGTAGCCTGAGAGAGCTGCTGCGACAGTCCACTACCTTTTTCAGAGTGAATCCC 60
Db 1 ATACGGGTAGCCTGAGAGAGCTGCTGCGACAGTCCACTACCTTTTTCAGAGTGAATCCC 60
Qy 61 GTTGTCCCAAGGCTTCCAGAGCGAACCTGTGCGGCTCGAGGACCGCGGCTCGAGTTT 120
Db 61 GTTGTCCCAAGGCTTCCAGAGCGAACCTGTGCGGCTCGAGGACCGCGGCTCGAGTTT 120
Qy 121 CGGCGTCCGGAAGACCGAGCTTCTTCGGGATCCAGTGTTCGTTTCAGGCCCCCAA 180
Db 121 CGGCGTCCGGAAGACCGAGCTTCTTCGGGATCCAGTGTTCGTTTCAGGCCCCCAA 180
Qy 181 TCTCAGCGCGGAGCGAG 214
Db 181 TCTCAGCGCGGAGCGAG 214

RESULT 2
BQ639435

LOCUS BQ639435 493 bp mRNA linear EST 15-JUL-2002
DEFINITION he1seil.y1 Human Retina cDNA (Un-normalized, unamplified): hd/he
Hom sapiens cDNA clone he1seil 5', mRNA sequence.

ACCESSION BQ639435
VERSION BQ639435.1 GI:21763894
KEYWORDS EST.

SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 493)
REFERENCE Wislow,G., Bernslein,S.L., Wyatt,M.K., Ray,S., Behal,A., Touchman
J.W., Bouffard,G., Smith,D. and Peterson,K.
Expressed sequence tag analysis of human retina for the NEIBank
Project: Retbindin, an abundant, novel retinal cDNA and alternative
splicing of other retina-preferred gene transcripts
Mol. Vis. 8 (4), (2002) In press
Contact: Wislow G

JOURNAL

COMMENT Section on Molecular Structure and Function
National Eye Institute
6/331, NIH, Bethesda, MD 20892-2740, USA

FEATURES
source

Location/Qualifiers
1. .493
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="he1seil"
/tissue_type="Retina"
/dev_stage="Adult"
/lab_host="EMDH108"
/clone_lib="Human Retina cDNA (Un-normalized, unamplified
): hd/he"

/note="Organ: Eye; Vector: pSPORT1; Neural retina tissue
was dissected from two 80 year old donors with no observed
eye disease. 100ug of total RNA was used for library
construction. A directionally cloned cDNA library in the
pSPORT1 vector (Life Technologies) was constructed at
Bioserve Biotechnology (Laurel MD) essentially following
the protocols of the SuperScript Plasmid System full
details of which are contained in the manufacturer's
instruction manual (http://www.lifetech.com/). First
strand synthesis was carried out using a Not I
primer-adaptor [5'-pGACTAGTCTAGATCGGAGCGGCCCT(T)15-3'
]. EST analysis was performed on the unamplified library
at the NIH Intramural Sequencing Center (NISC)."

BASE COUNT
ORIGIN

99 a 161 c 156 g 77 t
Query Match 98.8%; Score 212.4; DB 13; Length 493;
Best Local Similarity 99.5%; Pred. No. 7e-46;
Matches 213; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATACGGGTAGCCTGAGAGAGCTGCTGCGACAGTCCACTACCTTTTTCAGAGTGAATCCC 60
Db 10 ATACGGGTAGCCTGAGAGAGCTGCTGCGACAGTCCACTACCTTTTTCAGAGTGAATCCC 69
Qy 61 GTTGTCCCAAGGCTTCCAGAGCGAACCTGTGCGGCTCGAGGACCGCGGCTCGAGTTT 120
Db 70 GTTGTCCCAAGGCTTCCAGAGCGAACCTGTGCGGCTCGAGGACCGCGGCTCGAGTTT 129
Qy 121 CGGCGTCCGGAAGACCGAGCTTCTTCGGGATCCAGTGTTCGTTTCAGGCCCCCAA 180
Db 130 CGGCGTCCGGAAGACCGAGCTTCTTCGGGATCCAGTGTTCGTTTCAGGCCCCCAA 189
Qy 181 TCTCAGAGCGCGAGCGGAG 214
Db 190 TCTCAGAGCGGAGCGGAG 223

RESULT 3
CB154948

LOCUS CB154948 511 bp mRNA linear EST 29-JAN-2003
DEFINITION K-EST0213039 B2N807043 Homo sapiens cDNA clone B2N807043-30-G11 5',
mRNA sequence.

ACCESSION CB154948
VERSION CB154948.1 GI:28140051
KEYWORDS EST.

SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 511)
REFERENCE Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.

TITLE

JOURNAL 21C Frontier Korean EST Project 2001
COMMENT Unpublished
Contact: Kim YS
Genome Research Center

Tel: 301 402 3452
Fax: 301 496 0078
Email: Graeme@helix.nih.gov
Plate: 15 row: e column: 11
Seq primer: M13RP1 reverse primer (ABI).

Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 30 row: G column: 11
High quality sequence stop: 511.
Location/Qualifiers

FEATURES

source

1. 511
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="B2N807043-30-G11"
/sex="M"
/lab_host="Top10P"
/clone_lib="B2N807043"
/note="Organ: Brain; Vector: pcNS-D2; Site 1: EcoRI;
Site 2: NotI; The poly (A) + RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then decapped
with tobacco acid pyrophosphatase (TAP). The decapped
intact mRNA was ligated with DNA-RNA linker including
EcoRI site by treatment of T4 RNA ligase and the first
strand cDNA was synthesized from oligo dt-selected mRNA by
priming with dt-tailed vector. The dt-tailed vector was
adjusted to have about 60nt. The cDNA vector was
circularized with E. coli DNA ligase after digestion of
EcoRI which site is also included in vector. An RNA strand
converted to a DNA strand by Okayama-Berg method. The
obtained cDNA vectors were used for transformation of
competent cells E. coli Top10P by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library."

105 a 165 c 160 g 81 t
BASE COUNT
ORIGIN

Query Match 98.8%; Score 212.4; DB 14; Length 511;

Best Local Similarity 99.5%; Pred. No. 7.1e-46;

Matches 213; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATAACGGGTAGCTGAGGAGCTGCTGCGACAGTCCACTACCTTTTTCGAGAGTGACTCCC 60
DB 1 ATAACGGGTAGCTGAGGAGCTGCTGCGACAGTCCACTACCTTTTTCGAGAGTGACTCCC 60
QY 61 GTTGTCCCAAGCTTCCAGAGCGAACCTGTGCGGCTCGAGGACCGGCGGTCGAGTTT 120
DB 61 GTTGTCCCAAGCTTCCAGAGCGAACCTGTGCGGCTCGAGGACCGGCGGTCGAGTTT 120
QY 121 CCGGCTCCGGAAGGACCGAGCTCTTCTCGGGATCCAGTGTTCGTTCCAGCCCCCAA 180
DB 121 CCGGCTCCGGAAGGACCGAGCTCTTCTCGGGATCCAGTGTTCGTTCCAGCCCCCAA 180
QY 181 TCTCAGAGCGGAGCCGACAGAGCAGGGAACCG 214
DB 181 TCTCAGAGCGGAGCCGACAGAGCAGGGAACCG 214

RESULT 4

CB141518

LOCUS K-EST0195143 L15CKK1 Homo sapiens cDNA clone L15CKK1-28-A06 5',

DEFINITION mRNA sequence.

ACCESSION CB141518

VERSION CB141518.1 GI:28116735

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 527)

AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,

Oh,K.J., Cheong,W.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and

Kim,Y.S.

TITLE 21C Frontier Korean EST Project 2001

JOURNAL COMMENT

Unpublished

Contact: Kim YS

Genome Research Center

Korea Research Institute of Bioscience & Biotechnology

52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea

Tel: +82-42-860-4470

Fax: +82-42-860-4409

Email: yongsung@mail.kribb.re.kr

Plate: 28 row: A column: 06

High quality sequence stop: 527.

Location/Qualifiers

1. 527

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="L15CKK1-28-A06"

/sex="M"

/cell_line="CK-K1"

/lab_host="Top10P"

/clone_lib="L15CKK1"

/note="Organ: Liver; Vector: pcNS-D2; Site 1: EcoRI;
Site 2: NotI; The poly (A) + RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then decapped
with tobacco acid pyrophosphatase (TAP). The decapped
intact mRNA was ligated with DNA-RNA linker including
EcoRI site by treatment of T4 RNA ligase and the first
strand cDNA was synthesized from oligo dt-selected mRNA by
priming with dt-tailed vector. The dt-tailed vector was
adjusted to have about 60nt. The cDNA vector was
circularized with E. coli DNA ligase after digestion of
EcoRI which site is also included in vector. An RNA strand
converted to a DNA strand by Okayama-Berg method. The
obtained cDNA vectors were used for transformation of
competent cells E. coli Top10P by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library."

110 a 170 c 165 g 82 t

BASE COUNT

ORIGIN

Query Match 98.8%; Score 212.4; DB 14; Length 527;

Best Local Similarity 99.5%; Pred. No. 7.1e-46;

Matches 213; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATAACGGGTAGCTGAGGAGCTGCTGCGACAGTCCACTACCTTTTTCGAGAGTGACTCCC 60
DB 2 ATAACGGGTAGCTGAGGAGCTGCTGCGACAGTCCACTACCTTTTTCGAGAGTGACTCCC 61
QY 61 GTTGTCCCAAGCTTCCAGAGCGAACCTGTGCGGCTCGAGGACCGGCGGTCGAGTTT 120
DB 62 GTTGTCCCAAGCTTCCAGAGCGAACCTGTGCGGCTCGAGGACCGGCGGTCGAGTTT 121
QY 121 CCGGCTCCGGAAGGACCGAGCTCTTCTCGGGATCCAGTGTTCGTTCCAGCCCCCAA 180
DB 122 CCGGCTCCGGAAGGACCGAGCTCTTCTCGGGATCCAGTGTTCGTTCCAGCCCCCAA 181
QY 181 TCTCAGAGCGGAGCCGACAGAGCAGGGAACCG 214
DB 182 TCTCAGAGCGGAGCCGACAGAGCAGGGAACCG 215

RESULT 5

CB138426

LOCUS K-EST0191293 L15CKK1 Homo sapiens cDNA clone L15CKK1-29-E10 5',

DEFINITION mRNA sequence.

ACCESSION CB138426

VERSION CB138426.1 GI:28110126

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 590)

AUTHORS Kim, N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R., Oh, K.J., Cheong, J.E., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and Kim, Y.S.
TITLE 21C Frontier Korean EST Project 2001
JOURNAL Unpublished
COMMENT Contact: Kim YS
 Genome Research Center
 Korea Research Institute of Bioscience & Biotechnology
 52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
 Tel: +82-42-860-4470
 Fax: +82-42-860-4409
 Email: yongsung@mail.kribb.re.kr
 Plate: 29 row: E column: 10
 High quality sequence stop: 590.

FEATURES Location/Qualifiers
 1..590

source
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="L15CKK1-29-E10"
 /sex="M"
 /cell_line="CK-K1"
 /lab_host="Top10P"
 /clone_lib="L15CKK1"
 /note="Organ: Liver; Vector: pCNS-D2; Site 1: EcoRI; Site 2: NotI; The poly (A) + RNA was dephosphorylated with bacterial alkaline phosphatase (BAP) and then decapped with tobacco acid pyrophosphatase (TAP). The decapped intact mRNA was ligated with DNA-RNA linker including EcoRI site by treatment of T4 RNA ligase and the first strand cDNA was synthesized from oligo dt-selected mRNA by priming with dt-tailed vector. The dt-tailed vector was adjusted to have about 60nt. The cDNA vector was circularized with E. coli DNA ligase after digestion of EcoRI which site is also included in vector. An RNA strand converted to a DNA strand by Okayama-Berg method. The obtained cDNA vectors were used for transformation of competent cells E. coli Top10P by electroporation method. The cDNA libraries constructed by this method are full-length enriched cDNA library."
 126 a 186 c 185 g 93 t

BASE COUNT
ORIGIN

Query Match 98.8%; Score 212.4; DB 14; Length 590;
 Best Local Similarity 99.5%; Pred. No. 7.3e-46; Indels 0; Gaps 0;
 Matches 213; Conservative 0; Mismatches 1;
 QY 1 ATAACGGCTAGCTGAGGAGCTGCTGCGACAGTCCACTACCTTTTTCGAGAGTGACTCCC 60
 DB 1 ATAACGGCTAGCTGAGGAGCTGCTGCGACAGTCCACTACCTTTTTCGAGAGTGACTCCC 60
 QY 61 GTTGTCCCAAGGTTCCCAAGAGCAACCTGTGCGGTGAGGACCGCGCGCTCGAGTTT 120
 DB 61 GTTGTCCCAAGGTTCCCAAGAGCAACCTGTGCGGTGAGGACCGCGCGCTCGAGTTT 120
 QY 121 CCGGCGTCCGAGGAGGACCGAGCTCTTCGCGATCCAGTGTTCGTTTCCAGCCCCCAA 180
 DB 121 CCGGCGTCCGAGGAGGACCGAGCTCTTCGCGATCCAGTGTTCGTTTCCAGCCCCCAA 180
 QY 181 TCTCAGAGCGGAGCCGACAGAGCAGGGAACCG 214
 DB 181 TCTCAGAGCGGAGCCGACAGAGCAGGGAACCG 214

RESULT 6
LOCUS BG773197 701 bp mRNA linear EST 15-MAY-2001
DEFINITION BG773197 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4838721 5', mRNA sequence.
ACCESSION BG773197
VERSION BG773197.1 GI:14083850
KEYWORDS EST.
SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 701)

NIH-MGC http://mgc.nci.nih.gov/

Unpublished

REFERENCE Contact: Robert Strausberg, Ph.D.

AUTHORS Email: cgapbs-r@mail.nih.gov

TITLE Tissue Procurement: Miklos Palkovits, M.D., Ph.D.

JOURNAL CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki

COMMENT Toshiyuki and Piero Carninci (RIKEN)

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM10773 row: c column: 10

High quality sequence stop: 673.

FEATURES Location/Qualifiers

source

1..701

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:4838721"

/lab_host="DH10B"

/clone_lib="NIH_MGC_97"

/note="Organ: testis; Vector: pBluescriptR (modified

pBluescript KS+); Site1: BamHI, Site 2: SalI-XhoI (gtcgag

size-selected for average insert size 2.2 kb and

normalized to ROT 5. This is a primary library enriched

for full-length clones and constructed using the

Cap-trapper method (Carninci, in preparation). Library

constructed by M. Brownstein (NIH/NHGRI, National

Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT 153 a 227 c 209 g 112 t

ORIGIN

Query Match 98.8%; Score 212.4; DB 12; Length 701;

Best Local Similarity 99.5%; Pred. No. 7.5e-46; Indels 0; Gaps 0;

Matches 213; Conservative 0; Mismatches 1;

QY 1 ATAACGGCTAGCTGAGGAGCTGCTGCGACAGTCCACTACCTTTTTCGAGAGTGACTCCC 60

DB 370 ATAACGGCTAGCTGAGGAGCTGCTGCGACAGTCCACTACCTTTTTCGAGAGTGACTCCC 429

QY 61 GTTGTCCCAAGGTTCCCAAGAGCAACCTGTGCGGTGAGGACCGCGCGCTCGAGTTT 120

DB 430 GTTGTCCCAAGGTTCCCAAGAGCAACCTGTGCGGTGAGGACCGCGCGCTCGAGTTT 489

QY 121 CCGGCGTCCGAGGAGGACCGAGCTCTTCGCGATCCAGTGTTCGTTTCCAGCCCCCAA 180

DB 490 CCGGCGTCCGAGGAGGACCGAGCTCTTCGCGATCCAGTGTTCGTTTCCAGCCCCCAA 549

QY 181 TCTCAGAGCGGAGCCGACAGAGCAGGGAACCG 214

DB 550 TCTCAGAGCGGAGCCGACAGAGCAGGGAACCG 583

RESULT 7

LOCUS BG485554

DEFINITION BG485554

ACCESSION BG485554.1 GI:13417833

VERSION EST.

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 816)

816 bp mRNA linear EST 21-MAR-2001

602505445F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4618959 5',

mRNA sequence.

ACCESSION BG485554

VERSION BG485554.1 GI:13417833

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 816)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov
 Tissue Procurement: CLONTECH Laboratories, Inc.
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: LLM1377 row: j column: 12
 High quality sequence stop: 613.

FEATURES

source

```
1. .816
  /organism="Homo sapiens"
  /mol_type="mRNA"
  /db_xref="taxon:9606"
  /clone="IMAGE:4618859"
  /lab_host="DH10B (T1 phage-resistant)"
  /clone_lib="NIH MGC 77"
  /note="Organ: lung; Vector: pDNR-LIB (Clontech); Site: 1: SfiI (ggccgctggcc); Site 2: SfiI (ggccattggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCATTATGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCGCGCCGACATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.9 kb (range 0.5-4.0 kb). 12/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH MGC Library."
  BASE COUNT      183 a 261 c 263 g 109 t
  ORIGIN
```

Query Match 98.8%; Score 212.4; DB 10; Length 816;
 Best Local Similarity 99.5%; Pred. No. 7.7e-46;
 Matches 213; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 1 ATACGGCTAGCCTGAGGAGCTGTCGACAGTCCACTACCTTTTCGAGAGTGACTCCC 60
Db 2 ATACGGCTAGCCTGAGGAGCTGTCGACAGTCCACTACCTTTTCGAGAGTGACTCCC 61

QY 61 GTTGTCCTCCAGGCTTCCAGAGCGAACCTGTGCGGCTGAGGACCGCGCGCTCGAGTTT 120
Db 62 GTTGTCCTCCAGGCTTCCAGAGCGAACCTGTGCGGCTGAGGACCGCGCGCTCGAGTTT 121

QY 121 CCGGCGTCCGGAAGACCGAGCTCTTCTCGCGGATCCAGTGTTCGTTTCCAGCCCCCAA 180
Db 122 CCGGCGTCCGGAAGACCGAGCTCTTCTCGCGGATCCAGTGTTCGTTTCCAGCCCCCAA 181

QY 181 TCTCAGAGCGGAGCCGACAGAGCAGGGAACCG 214
Db 182 TCTCAGAGCGGAGCCGACAGAGCAGGGAACCG 215
```

RESULT 8
BI462679
LOCUS 603202121F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5268172 5',
DEFINITION mRNA sequence.

ACCESSION BI462679
VERSION BI462679.1 GI:15253335
KEYWORDS EST.
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 916)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: LLM11676 row: i column: 05
 High quality sequence stop: 703.

FEATURES

source

```
1. .916
  /organism="Homo sapiens"
  /mol_type="mRNA"
  /db_xref="taxon:9606"
  /clone="IMAGE:5268172"
  /lab_host="DH10B"
  /clone_lib="NIH MGC 97"
  /note="Organ: testis; Vector: pBluescriptR (modified pBluescript KS+); Site 1: BamHI; Site 2: Sall-XhoI (gtcgag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.2 kb and normalized to 500. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: this is a NIH MGC Library."
  BASE COUNT      196 a 289 c 293 g 138 t
  ORIGIN
```

Query Match 98.8%; Score 212.4; DB 12; Length 916;
 Best Local Similarity 99.5%; Pred. No. 7.8e-46;
 Matches 213; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 1 ATACGGCTAGCCTGAGGAGCTGTCGACAGTCCACTACCTTTTCGAGAGTGACTCCC 60
Db 8 ATACGGCTAGCCTGAGGAGCTGTCGACAGTCCACTACCTTTTCGAGAGTGACTCCC 67

QY 61 GTTGTCCTCCAGGCTTCCAGAGCGAACCTGTGCGGCTGAGGACCGCGCGCTCGAGTTT 120
Db 68 GTTGTCCTCCAGGCTTCCAGAGCGAACCTGTGCGGCTGAGGACCGCGCGCTCGAGTTT 127

QY 121 CCGGCGTCCGGAAGACCGAGCTCTTCTCGCGGATCCAGTGTTCGTTTCCAGCCCCCAA 180
Db 128 CCGGCGTCCGGAAGACCGAGCTCTTCTCGCGGATCCAGTGTTCGTTTCCAGCCCCCAA 187

QY 181 TCTCAGAGCGGAGCCGACAGAGCAGGGAACCG 214
Db 188 TCTCAGAGCGGAGCCGACAGAGCAGGGAACCG 221
```

RESULT 9
BI464037
LOCUS 603202811F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5268482 5',
DEFINITION mRNA sequence.

ACCESSION BI464037
VERSION BI464037.1 GI:15254693
KEYWORDS EST.
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 751)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN)

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLM11677 row: f column: 03
 High quality sequence stop: 740.

FEATURES

Location/Qualifiers

1..751
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5268482"
 /lab_host="DH10B"
 /clone_lib="NIH MGC 97"
 /note="Organ: testis; Vector: pBluescriptR (modified
 pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (Gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
 size-selected for average insert size 2.2 kb and
 normalized to ROT 5. This is a primary library enriched
 for full-length clones and constructed using the
 Cap-trapper method (Carninci, in preparation). Library
 constructed by M. Brownstein (NIMH/NHGRI, National
 Institutes of Health). Note: this is a NIH_MGC Library."
 162 a 233 c 241 g 115 t

BASE COUNT

ORIGIN

Query Match 98.3%; Score 211.4; DB 12; Length 751;
 Best Local Similarity 99.5%; Pred. No. 1.4e-45;
 Matches 212; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 2 TAACGGCTAGCTGAGGAGCTCTCGACAGTCCACTACCTTTTCGAGAGTGACTCCCG 61
 Db 9 TAACGGCTAGCTGAGGAGCTCTCGACAGTCCACTACCTTTTCGAGAGTGACTCCCG 68
 QY 62 TTGTCCCAAGGTTCCCGAGCGAACCCTGTGCGGCTGCAGGACCGCGCGCTCGAGTTTC 121
 Db 69 TTGTCCCAAGGTTCCCGAGCGAACCCTGTGCGGCTGCAGGACCGCGCGCTCGAGTTTC 128
 QY 122 CGCGCTCGGAAGGACCGAGCTTCTTCGCGGATCCAGTGTTCCTCCAGCCCCCAAT 181
 Db 129 CGCGCTCGGAAGGACCGAGCTTCTTCGCGGATCCAGTGTTCCTCCAGCCCCCAAT 188
 QY 182 CTCAGAGCGGACCGACAGAGGAGGAAACCG 214
 Db 189 CTCAGAGCGGACCGACAGAGGAGGAAACCG 221

RESULT 10

AU076824
 LOCUS 267 bp mRNA linear EST 04-MAY-2000
 DEFINITION AU076824 Sugano cDNA library Homo sapiens cDNA clone kaial078
 similar to 5'-end region of Human MHC class III HSP70-1 gene (HLA),
 mRNA sequence.

ACCESSION

AU076824
 VERSION AU076824.1 GI:7439327

KEYWORDS

EST.
 SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 267)
 AUTHORS Suzuki.Y., Ishihara.D., Sasaki.M., Nakagawa.H., Hata.H., Tsunoda.T.,
 Watanabe.M., Komatsu.T., Ota.T., Isogai.T., Suyama.A. and Sugano
 S.

TITLE

Statistical analysis of the 5' untranslated region of human mRNA
 using 'Oligo-Capped' cDNA libraries

JOURNAL

Genomics 64 (3), 286-297 (2000)

MEDLINE

20221373

PUBMED

10756036

COMMENT

Contact: Yutaka Suzuki
 Department of Virology
 Institute of Medical Science, University of Tokyo

4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
 Email: yusuzuki@ims.u-tokyo.ac.jp
 Suzuki.Y., Yoshitomo-Nakagawa.K., Maruyama.K., Suyama.A. and Sugano
 S. Construction and characterization of a full length-enriched and
 a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997)
 This clone was obtained from a 'full length-enriched' cDNA library
 constructed by 'Oligo-Capping' method. The coding region starts
 from the 50 bp upstream to the 3'-end.

FEATURES

Location/Qualifiers

1..267
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="kaial078"
 /clone_lib="Sugano cDNA library"
 51 a 85 c 77 g 48 t 2 others

BASE COUNT

ORIGIN

Query Match 97.9%; Score 210.4; DB 9; Length 267;
 Best Local Similarity 98.6%; Pred. No. 2.1e-45;
 Matches 211; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 ATAACGGCTAGCTGAGGAGCTGTGCGACAGTCCACTACCTTTTCGAGAGTGACTCCC 60
 Db 2 ATAACGGCTAGCTGAGGAGCTGTGCGACAGTCCACTACCTTTTCGAGAGTGACTCCC 61
 QY 61 GTTGTCCCAAGGTTCCCGAGCGAACCCTGTGCGGCTGCAGGACCGCGCGCTCGAGTTT 120
 Db 62 GTTGTCCCAAGGTTCCCGAGCGAACCCTGTGCGGCTGCAGGACCGCGCGCTCGAGTTT 121
 QY 121 CCGCGCTCCGGAAGGACCGAGCTCTTTCGCGGATCCAGTGTTCCTCCAGCCCCCAA 180
 Db 122 CCGCGCTCCGGAAGGACCGAGCTCTTTCGCGGATCCAGTGTTCCTCCAGCCCCCAA 181
 QY 181 TCTCAGAGCGGACCGACAGAGGAGGAAACCG 214
 Db 182 TCTCAGAGCGGACCGACAGAGGAGGAAACCG 215

RESULT 11

BM822968
 LOCUS 316 bp mRNA linear EST 06-MAR-2002
 DEFINITION K-EST0093271 S21SNUS20s1 Homo sapiens cDNA clone S21SNUS20s1-10-C06
 5', mRNA sequence.

ACCESSION

BM822968
 VERSION BM822968.1 GI:19179381

KEYWORDS

EST.
 SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 316)
 AUTHORS Kim,N.S., Hahn,X., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
 Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
 Kim,Y.S.

TITLE

21C Frontier Korean EST Project 2001

JOURNAL

Unpublished

COMMENT

Contact: Kim YS
 Genome Research Center
 Korea Research Institute of Bioscience & Biotechnology
 52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
 Tel: +82-42-860-4470
 Fax: +82-42-860-4409
 Email: yongsung@mail.kribb.re.kr
 Plate: 10 row: C column: 06
 High quality sequence stop: 316.
 Location/Qualifiers

FEATURES

1..316
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="S21SNUS20s1-10-C06"
 /sex="F"

```

/tissue_type="Stomach"
/cell_type="Floating aggregates"
/cell_line="SNU-520"
/lab_host="Top10P"
/clone_lib="S21SNU520s1"
/notes="Organ: Stomach; Vector: pTZ19RP1; Site_1: EcoRI;
Site_2: NotI; The poly (A)+ RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then decapped
with tobacco acid pyrophosphatase (TAP). The decapped
intact mRNA was ligated with DNA-RNA linker including EcoR
I site by treatment of T4 RNA ligase and the first strand
cDNA was synthesized from oligo dt-selected mRNA by
priming with dt-tailed vector. The dt-tailed vector was
adjusted to have about 60nt. The cDNA vector was
circularized with E. coli DNA ligase after digestion of
EcoRI which site is also included in vector. An RNA strand
obtained cDNA vectors were used for transformation of
competent cells E. coli Top10P by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library. After analyzing and
sequencing about 2,000 - 3,000 colonies in original cDNA
library, the abundant cDNAs were selected and amplified by
PCR reaction using vector region primer including T7
promotor as 5' primer and N(dT)14 as 3' primer. The PCR
products were used as template for synthesis of
biotinylated single stranded RNA by in vitro transcription
reaction. The synthesized RNA probes were hybridized with
antisense single stranded cDNAs prepared from original
library and incubated with avidin-gel. After removing
DNA-RNA hybrids by centrifuge, the subtracted cDNA
libraries were constructed by transformation of the
remaining DNA into competent cells E. coli Top10P with
electroporation method."
BASE COUNT      62 a 103 c 96 g
ORIGIN
.. Query Match      97.9%; Score 210.4; DB 12; Length 316;
Best Local Similarity 99.5%; Pred. No. 2.2e-45;
Matches 211; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AACGCTAGCTGAGAGCTGTCGACAGTCCACTACTCTTTTCGAGGTGACTCCCGT 62
DB 6 AACGCTAGCTGAGAGCTGTCGACAGTCCACTACTCTTTTCGAGGTGACTCCCGT 65

QY 63 TGTCCCAAGGCTTCCAGAGCAACTGTGCGGTGCGAGCACCGCGGTGAGTTTCC 122
DB 66 TGTCCCAAGGCTTCCAGAGCAACTGTGCGGTGCGAGCACCGCGGTGAGTTTCC 125

QY 123 GCGGTCCGAGAGCCGAGCTCTTCTCGCGGATCCAGTGTTCGGTTTCAGCCCCCAATC 182
DB 126 GCGGTCCGAGAGCCGAGCTCTTCTCGCGGATCCAGTGTTCGGTTTCAGCCCCCAATC 185

QY 183 TCAGAGCCGAGCCGACAGAGCAGGGAACCG 214
DB 186 TCAGAGCCGAGCCGACAGAGCAGGGAACCG 217

```

```

RESULT 12
CB133590
LOCUS
DEFINITION
K-EST0184602 L4SNU368s1 Homo sapiens cDNA clone L4SNU368s1-7-H08
5', mRNA sequence.
ACCESSION
CB133590
VERSION
CB133590.1 GI:28099354
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 376)
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and

```

TITLE JOURNAL COMMENT

Kim,Y.S.
21C Frontier Korean EST Project 2001
Unpublished
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 7 row: H column: 08
High quality sequence stop: 376.
Location/Qualifiers
1. 376

FEATURES source

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="L4SNU368s1-7-H08"
/sex="M"
/tissue_type="Liver"
/cell_type="Polygonal"
/cell_line="SNU-368"
/lab_host="Top10P"
/clone_lib="L4SNU368s1"
/note="Organ: Liver; Vector: pCNS-D2; Site_1: EcoRI;
Site_2: NotI; The poly (A)+ RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then decapped
with tobacco acid pyrophosphatase (TAP). The decapped
intact mRNA was ligated with DNA-RNA linker including
EcoRI site by treatment of T4 RNA ligase and the first
strand cDNA was synthesized from oligo dt-selected mRNA by
priming with dt-tailed vector. The dt-tailed vector was
adjusted to have about 60nt. The cDNA vector was
circularized with E. coli DNA ligase after digestion of
EcoRI which site is also included in vector. An RNA strand
converted to a DNA strand by Okayama-Berg method. The
obtained cDNA vectors were used for transformation of
competent cells E. coli Top10P by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library. After analyzing and
sequencing about 2,000 - 3,000 colonies in original cDNA
library, the abundant cDNAs were selected and amplified by
PCR reaction using vector region primer including T7
promotor as 5' primer and N(dT)14 as 3' primer. The PCR
products were used as template for synthesis of
biotinylated single stranded RNA by in vitro transcription
reaction. The synthesized RNA probes were hybridized with
antisense single stranded cDNAs prepared from original
library and incubated with avidin-gel. After removing
DNA-RNA hybrids by centrifuge, the subtracted cDNA
libraries were constructed by transformation of the
remaining DNA into competent cells E. coli Top10P with
electroporation method."

BASE COUNT
ORIGIN

Query Match 97.9%; Score 210.4; DB 14; Length 376;
Best Local Similarity 99.5%; Pred. No. 2.3e-51;
Matches 211; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AACGCTAGCTGAGAGCTGTCGACAGTCCACTACTCTTTTCGAGGTGACTCCCGT 62
DB 1 AACGCTAGCTGAGAGCTGTCGACAGTCCACTACTCTTTTCGAGGTGACTCCCGT 60

QY 63 TGTCCCAAGGCTTCCAGAGCAACTGTGCGGTGCGAGCACCGCGGTGAGTTTCC 122
DB 61 TGTCCCAAGGCTTCCAGAGCAACTGTGCGGTGCGAGCACCGCGGTGAGTTTCC 120

QY 123 GCGGTCCGAGAGCCGAGCTCTTCTCGCGGATCCAGTGTTCGGTTTCAGCCCCCAATC 182
DB 121 GCGGTCCGAGAGCCGAGCTCTTCTCGCGGATCCAGTGTTCGGTTTCAGCCCCCAATC 180

QY 183 TCAGAGCCGAGCCGACAGAGCAGGGAACCG 214

```

Db      181 TCAGAGCGGACCGACAGAGCGGGAACCG 212
|||||
CB133685      409 bp mRNA linear EST 29-JAN-2003
K-BST0184769 L4SNU36861 Homo sapiens cDNA clone L4SNU36861-19-G12
5', mRNA sequence.
CB133685
CB133685      1 GI:28099498
EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS      Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
TITLE      21C Frontier Korean EST Project 2001
JOURNAL      Unpublished
COMMENT      Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 19 row: G column: 12
High quality sequence stop: 409.
Location/Qualifiers
1..409
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="L4SNU36861-19-G12"
/sex="M"
/tissue_type="Liver"
/cell_type="Polygonal"
/lab_host="SNU-368"
/clone_lib="L4SNU36861"
/note="Organ: Liver; Vector: PCNS-D2; Site 1: EcoRI;
Site 2: NotI; The poly (A)+ RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then decapped
with tobacco acid pyrophosphatase (TAP). The decapped
intact mRNA was ligated with DNA-RNA linker including
EcoRI site by treatment of T4 RNA ligase and the first
strand cDNA was synthesized from oligo dt-selected mRNA by
priming with dt-tailed vector. The dt-tailed vector was
adjusted to have about 60nt. The cDNA vector was
circularized with E. coli DNA ligase after digestion of
EcoRI which site is also included in vector. An RNA strand
converted to a DNA strand by Okayama-Berg method. The
obtained cDNA vectors were used for transformation of
competent cells E. coli Top10F by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library. After analyzing and
sequencing about 2,000 - 3,000 colonies in original cDNA
library, the abundant cDNAs were selected and amplified by
PCR reaction using vector region primer including T7
promotor as 5' primer and N(dT)14 as 3' primer. The PCR
products were used as template for synthesis of
biotinylated single stranded RNA by in vitro transcription
reaction. The synthesized RNA probes were hybridized with
antisense single stranded cDNAs prepared from original
library and incubated with avidin-gel. After removing
DNA-RNA hybrids by centrifuge, the subtracted cDNA
libraries were constructed by transformation of the
remaining DNA into competent cells E. coli Top10F, with
electroporation method."
BASE COUNT      85 a 137 c 124 g 63 t

ORIGIN
Query Match      97.9%; Score 210.4; DB 14; Length 409;
Best Local Similarity 99.5%; Pred. No. 2.3e-45;
Matches 211; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3 AACGGCTAGCCTGAGGAGCTGCTCGACAGTCCACTACCTTTTTCGAGAGTGACTCCCGT 62
|||||
DB      1 AACGGCTAGCCTGAGGAGCTGCTCGACAGTCCACTACCTTTTTCGAGAGTGACTCCCGT 60
|||||

QY      63 TGTCCCAAGGCTTCCACAGAGCAACCTGTGCGGTGACGACCGCGCGCTCGAGTTTCC 122
|||||
DB      61 TGTCCCAAGGCTTCCACAGAGCAACCTGTGCGGTGACGACCGCGCGCTCGAGTTTCC 120
|||||

QY      123 GGCGTCCGAGGAGCGGAGCTCTTCTCGCGATCCAGTGTTCGTTTCCAGCCCCCAATC 182
|||||
DB      121 GGCGTCCGAGGAGCGGAGCTCTTCTCGCGATCCAGTGTTCGTTTCCAGCCCCCAATC 180
|||||

QY      183 TCAGAGCGGACCGACAGAGCGGGAACCG 214
|||||
DB      181 TCAGAGCGGACCGACAGAGCGGGAACCG 212
|||||

RESULT 14
BI548281
LOCUS      603191148P1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5262587 5',
DEFINITION      mRNA sequence.
ACCESSION      BI548281
VERSION      BI548281.1 GI:15435593
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 436)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-rc@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM11661 row: p column: 12
High quality sequence stop: 431.
Location/Qualifiers
1..436
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5262587"
/tissue_type="hippocampus"
/lab_host="DH10B"
/clone_lib="NIH_MGC_95"
/note="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TGTGTTTGTGTTTGTGTTT-3',
size-selected for average insert size 2.5 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT      90 a 143 c 134 g 67 t 2 Others

ORIGIN
Query Match      97.9%; Score 210.4; DB 12; Length 436;

```

Best Local Similarity 99.5%; Pred. No. 2.3e-45;
Matches 211; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AACGGCTAGCTGAGGAGCTGCTGGACAGTCCACTTCTTTTCGAGAGTCACTCCCGT 62
Db 6 AACGGCTAGCTGAGGAGCTGCTGGACAGTCCACTTCTTTTCGAGAGTCACTCCCGT 65

QY 63 TGTCCCAAGGCTTCCAGAGCAACTGTGCGGCTGCAGGCAACCGCGGTGAGTTTCC 122
Db 66 TGTCCCAAGGCTTCCAGAGCAACTGTGCGGCTGCAGGCAACCGCGGTGAGTTTCC 125

QY 123 GCGCTCCGAAGACGAGCTCTTCTCGCGGATCAGTGTTCGTTTCAGGCCCCAATC 182
Db 126 GCGCTCCGAAGACGAGCTCTTCTCGCGGATCAGTGTTCGTTTCAGGCCCCAATC 185

QY 183 TCAGAGCGGAGCGGACAGAGCAGGGAACCG 214
Db 186 TCAGAGCGGAGCGGACAGAGCAGGGAACCG 217

RESULT 15
CB137170
LOCUS
DEFINITION CB137170 463 bp mRNA linear EST 29-JAN-2003
K-EST0189826 L4SNU368s1 Homo sapiens cDNA clone L4SNU368s1-21-B08
5', mRNA sequence.
ACCESSION CB137170
VERSION CB137170.1 GI:28105049
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 463)
Kim, N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R.,
Oh, K.J., Cheong, J.E., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and
Kim, Y.S.

TITLE 21C Fronttier Korean EST Project 2001

JOURNAL Unpublished
COMMENT Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 21 row: B column: 08
High quality sequence stop: 463.

FEATURES
source
1..463
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="L4SNU368s1-21-B08"
/sex="M"
/tissue_type="Liver"
/cell_type="polygonal"
/cell_line="SNU-368"
/lab_host="Top10P"
/clone_lib="L4SNU368s1"
/note="Organ: Liver; Vector: pCNS-D2; Site 1: EcoRI;
Site 2: NotI; The poly (A)+ RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then decapped
with tabacco acid pyrophosphatase (TAP). The decapped
intact mRNA was ligated with DNA-RNA linker including
EcoRI site by treatment of T4 RNA ligase and the first
strand cDNA was synthesized from oligo dt-selected mRNA by
priming with dt-tailed vector. The dt-tailed vector was
adjusted to have about 60nt. The cDNA vector was
circularized with E. coli DNA ligase after digestion of
EcoRI which site is also included in vector. An RNA strand
converted to a DNA strand by Okayama-Berg method. The
obtained cDNA vectors were used for transformation of
competent cells E. coli Top10P by electroporation method.

The cDNA libraries constructed by this method are full-length enriched cDNA library. After analyzing and sequencing about 2,000 - 3,000 colonies in original cDNA library, the abundant cDNAs were selected and amplified by PCR reaction using vector region primer including T7 promoter as 5' primer and N(dT)14 as 3' primer. The PCR products were used as template for synthesis of biotinylated single stranded RNA by in vitro transcription reaction. The synthesized RNA probes were hybridized with antisense single stranded cDNAs prepared from original library and incubated with avidin-gel. After removing DNA-RNA hybrids by centrifuge, the subtracted cDNA libraries were constructed by transformation of the remaining DNA into competent cells E. coli Top10P with electroporation method."

BASE COUNT 92 a 153 c 146 g 72 t
ORIGIN

Query Match 97.9%; Score 210.4; DB 14; Length 463;
Best Local Similarity 99.5%; Pred. No. 2.3e-45;
Matches 211; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AACGGCTAGCTGAGGAGCTGCTGGACAGTCCACTTCTTTTCGAGAGTCACTCCCGT 62
Db 1 AACGGCTAGCTGAGGAGCTGCTGGACAGTCCACTTCTTTTCGAGAGTCACTCCCGT 60

QY 63 TGTCCCAAGGCTTCCAGAGCAACTGTGCGGCTGCAGGCAACCGCGGTGAGTTTCC 122
Db 61 TGTCCCAAGGCTTCCAGAGCAACTGTGCGGCTGCAGGCAACCGCGGTGAGTTTCC 120

QY 123 GCGCTCCGAAGACGAGCTCTTCTCGCGGATCAGTGTTCGTTTCAGGCCCCAATC 182
Db 121 GCGCTCCGAAGACGAGCTCTTCTCGCGGATCAGTGTTCGTTTCAGGCCCCAATC 180

QY 183 TCAGAGCGGAGCGGACAGAGCAGGGAACCG 214
Db 181 TCAGAGCGGAGCGGACAGAGCAGGGAACCG 212

Search completed: February 3, 2004, 03:22:16
Job time : 1231 secs

THIS PAGE BLANK (USPTO)